HAMK
Häme University of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin osarahoittama

Introduction to Machine Learning

The workshop is part of EU co-funded Robo CO. - Robots and AI as future colleagues –project

HAMK
Häme University
of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin
osarahoittama

Anastasiia Stepanova
Project engineer
HAMK Tech Robotics
anastasiia.stepanova@hamk.fi

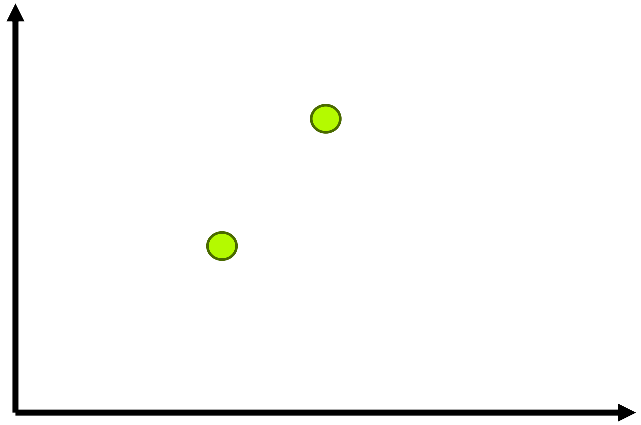
Euroopan unionin
osarahoittama

What is AI?

- AI refers to the development of computer systems that can perform tasks requiring human intelligence, such as visual perception, speech recognition, decision-making, and language translation.
- AI involves the creation of algorithms and models that enable machines to learn from data and make intelligent decisions.

Some core ideas to understand ML

1. Engineering approach (problem of points in the space).
2. CS → algorithmic approach:
   - Formulate the points in the space into another data structure: your favorite data structure?

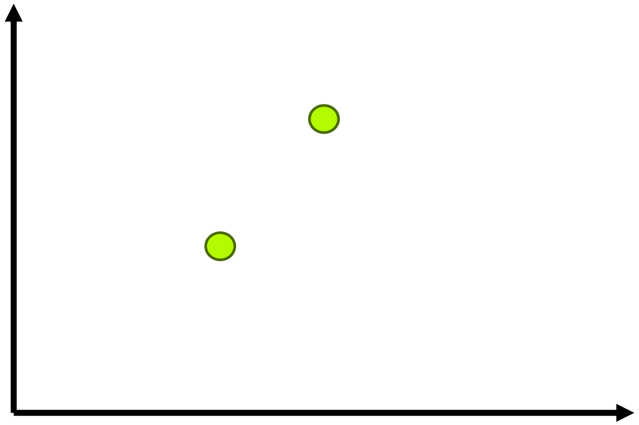


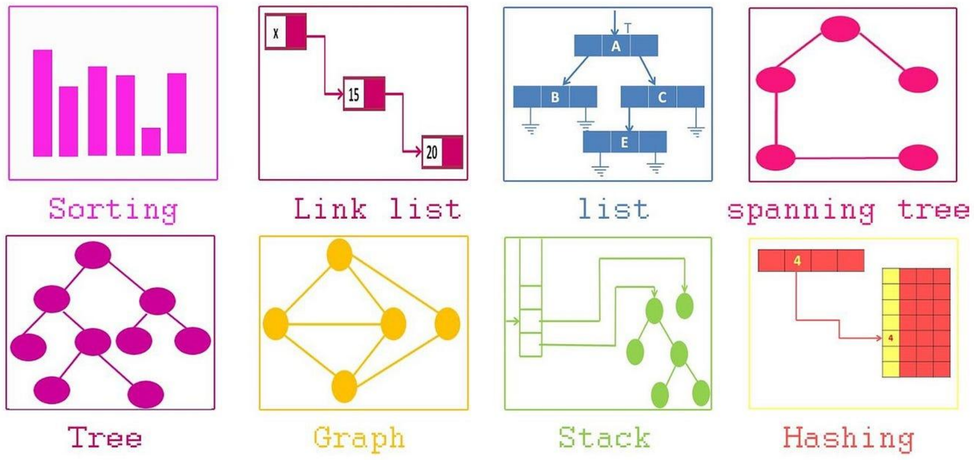
Some core ideas to understand ML

1. Engineering approach (problem of points in the space).

2. CS -> algorithmic approach:

- Formulate the points in the space into another data structure: your favorite data structure? (list, graph, decision tree…)





HAMK
Häme University
of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin
osarahoittama

Starting with engineering approach

Linear regression

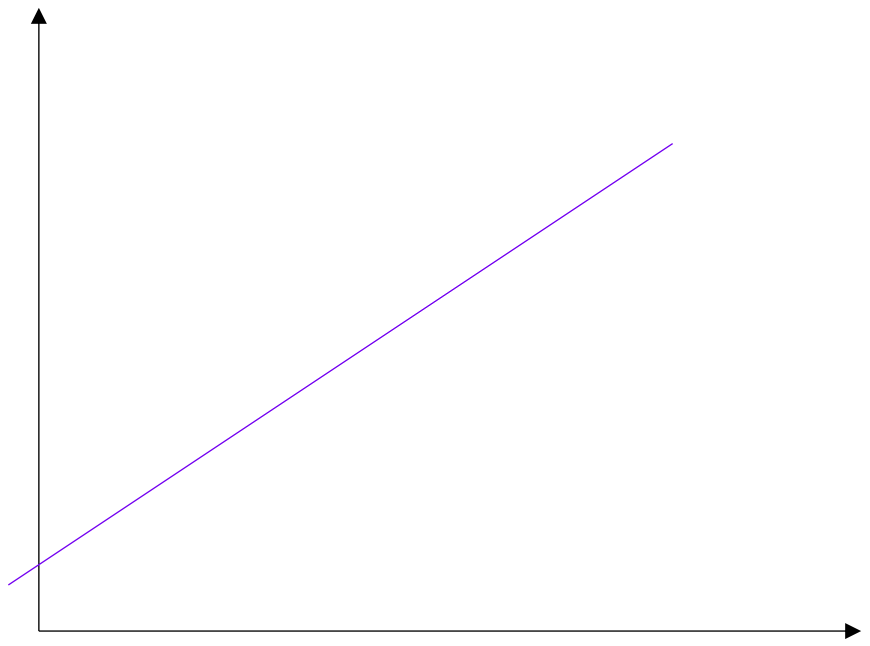
- Nothing more complicated than math from high school ☺

Linear regression

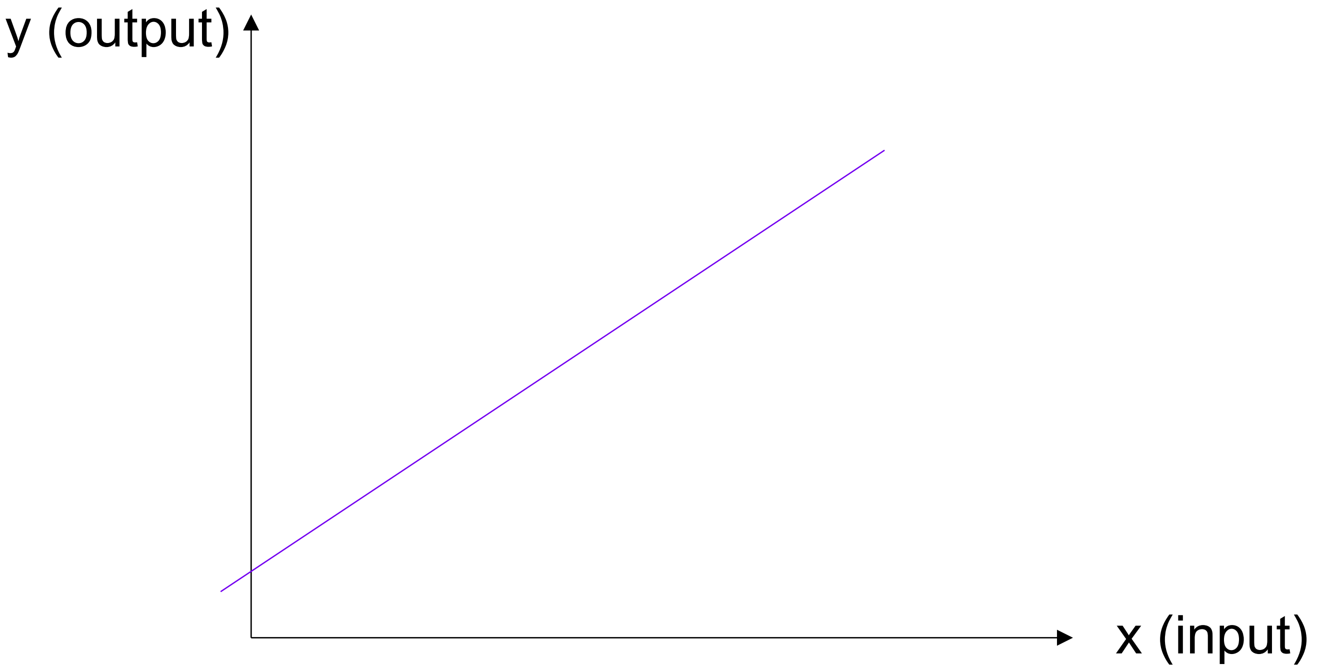
- Nothing more complicated than math from high school ☺
- Formulas of linear regression from high school:

  y = ax + b -> this is f(x) – some function

  This simple formula is behind LLMs!



- $y = ax + b$ → Maps observation **x** to desired output **y** (observation can be an image)



HAMK
Häme University
of Applied Sciences

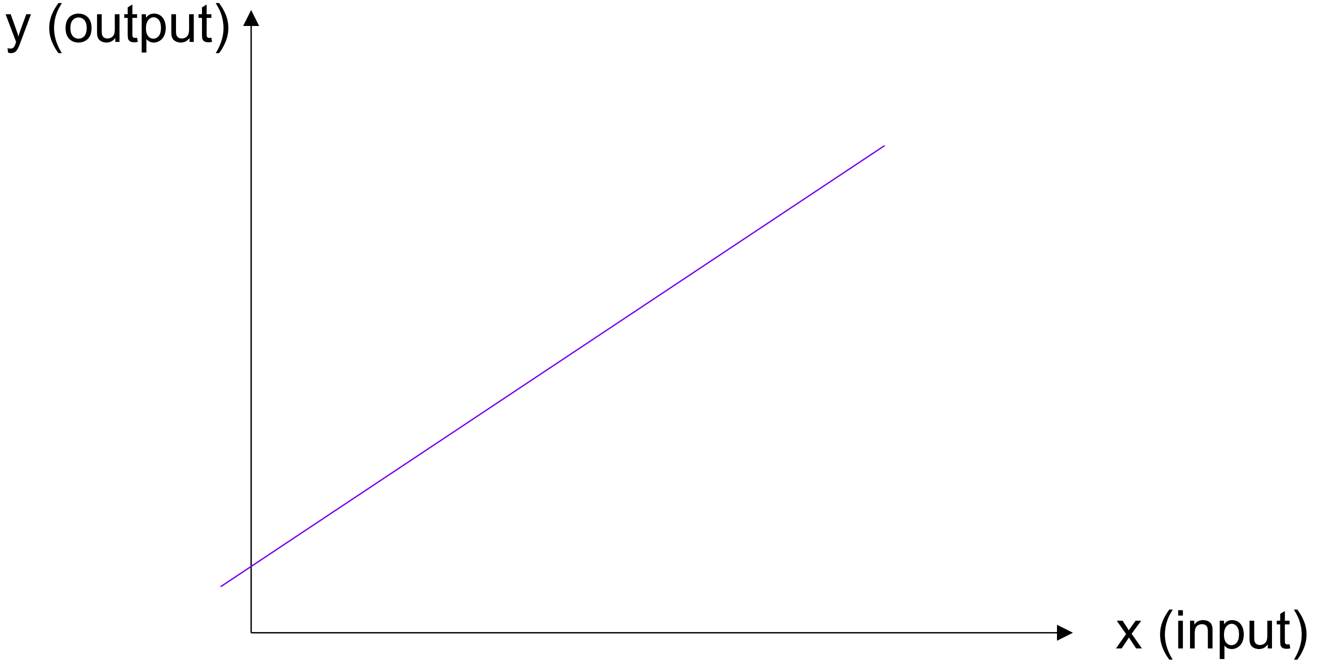
REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin
osarahoittama

- $y = \underbrace{a}_{\text{slope}}x + \underbrace{b}_{\text{intercept}}$

Maps observation x to desired output y
(observation can be an image)

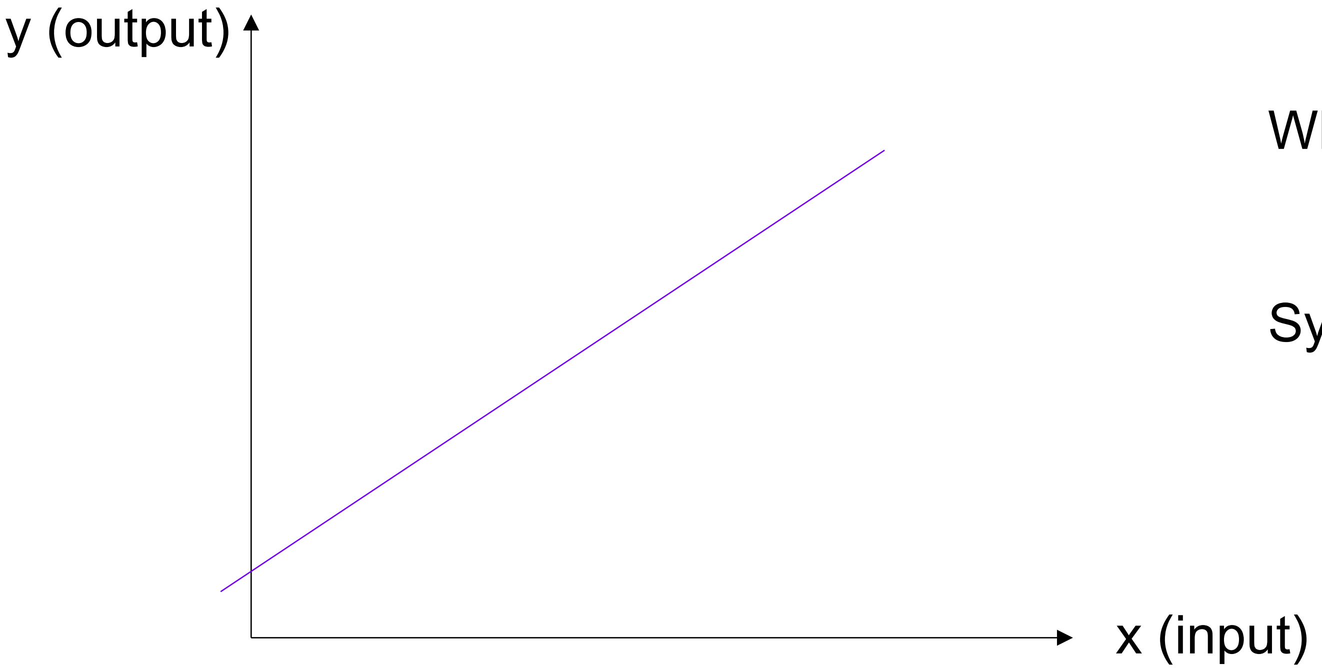
Example
Face picture: age, stress level…



- $y = ax + b$ → Maps observation x to desired output y (observation can be an image)

slope intercept

Example
Face picture: age, stress level…



What is learning here?

⇩

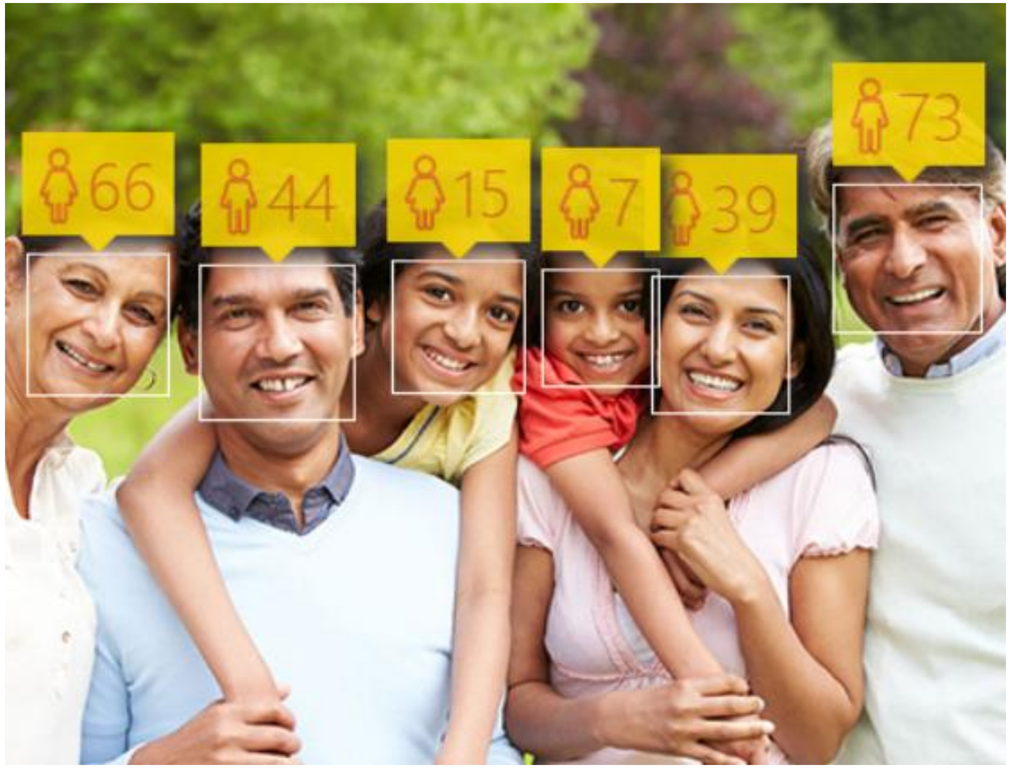
System learns → to values a and b

HAMK Häme University of Applied Sciences

RUN Regional University Network European University

Euroopan unionin osarahoittama

Age estimation on image



Example

- Let’s assume a Hobbit MD, who has money to buy only one new instrument: either a scale for measuring weight or ruler to measure height.
- Hobbit MD wants to save money, and he is ML enthusiast, so he thinks that he can estimate weight y if he buys only a cheap measuring tape for height.
- Simply measure height x and use some ML model to estimate weight $y = f(x)$.
- Here, we know that height and weight are correlated (taller hobbits tend to be heavier).

Example

- Let's assume a Hobbit MD, who has money to buy only one new instrument: either a scale for measuring weight or ruler to measure height.
- Hobbit MD wants to save money, and he is ML enthusiast, so he thinks that he can estimate weight y if he buys only a cheap measuring tape for height.
- Simply measure height x and use some ML model to estimate weight $y = f(x)$.
- Here, we know that height and weight are correlated (taller hobbits tend to be heavier).
- → Fit the model

HAMK
Häme University
of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin
osarahoittama

Back to high school

- **y = ax + b**, what are **a** and **b** called here?

Back to high school

- **y = ax + b**

slope

Intercept or bias

(Tells the interception point
where it hits y-axis)

→ Machine Learning approach to the problem

ML approach

- Start with training data:
  - Measured example and we know the output value for it

x1 → y1
x2 → y2
x3 → y3
….

Learn from the training data and adjust **a** and **b** values.

ML approach

- Start with training data:
  - Measured example and we know output value for it

x1 → y1
x2 → y2
x3 → y3
….

- Learn from the training data and adjust **a** and **b** values.

- **How Hobbit gets the training data?**
  → Measure other hobbits' height and scale, ask permission to use the data to train ML model.

HAMK
Häme University
of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin
osarahoittama

What if zero samples?

- Then every line is equally good for us.
- Randomly pick **a** and **b**.

If we have training data

- **We have only 1 sample**
  - (x1,y1) = (1.11 m, 85.0 kg) -> **gives us one point**
  - Solve it: y = ax +b
    - 85 = 1.11*a +b
    - b = 85 – 1.11*a

Now all lines will go through the same point!

HAMK
Häme University of Applied Sciences

Euroopan unionin osarahoittama

Continue training data

- **We have 2 samples**
  - (x2,y2) = (1.52 m, 110 kg)
  - 110 = 1.52*a + b

Now with two points we can draw a line!

  - From sample 1: b = 85 – 1.11*a
  - Substitute to sample 2: 110 = 1.52*a + 85 – 1.11*a
  - **a** = (110 – 85)/(1.52-1.11)
  - Solve **b**: b = 85 – 1.11*a = …
  - Get both **a** and **b** and draw the line

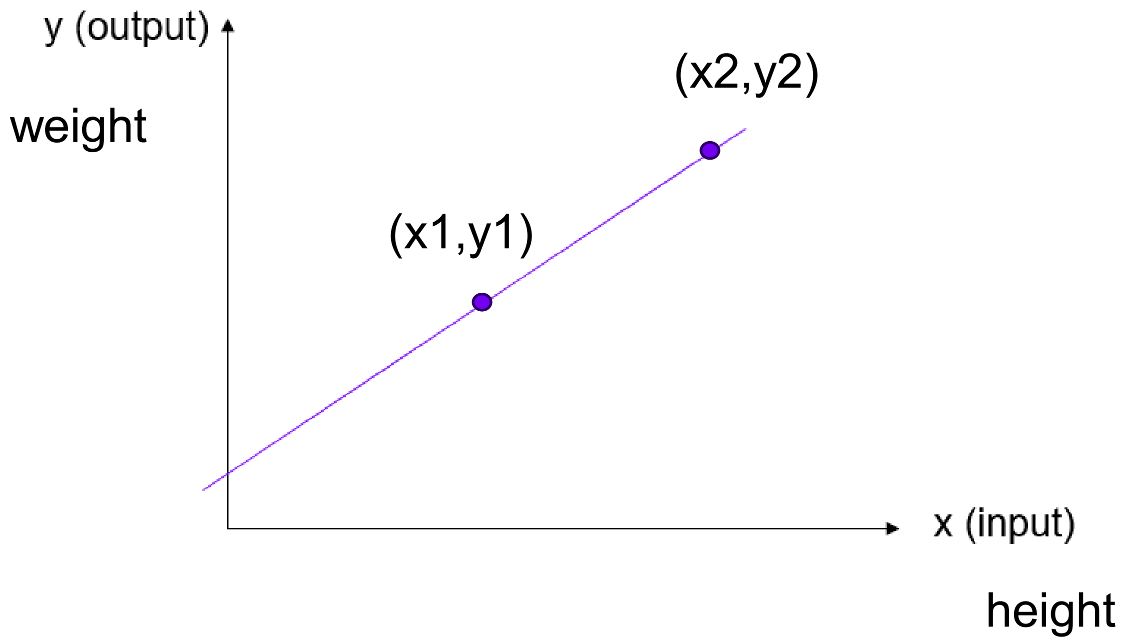
HAMK
Häme University
of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin
osarahoittama

Linear regression

- **y = ax + b**
- Two points → line through them

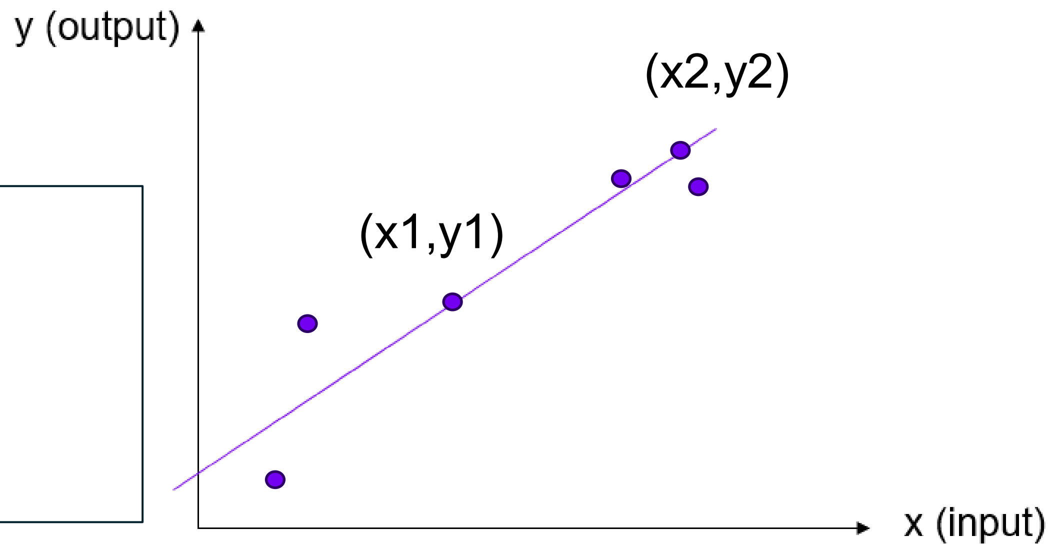


Linear regression

- **y = ax + b**
- Now add more points
  - The more points, the better

x1 → y1
x2 → y2
x3 → y3
....

Training data



Some problem: fit the line to multiple points

How do we use training data now?

- Take new hobbit as test sample

- What distinguishes training samples and test samples?

Training sample vs test sample

- **Training samples:** used to train the model (model adjusts its parameters) based on the training data to minimize error and improve performance.
  - The model has access to both the input features and the corresponding labels or outputs.
  - Train samples help the model generalize to unseen data by learning from examples.

➢ **Test samples:** used to evaluate the model's performance on unseen data, ensuring that the model has learned to generalize and not just memorize the training data.

  ➢ The model uses the input features of the test data to make predictions, which are then compared to the actual labels to assess its accuracy.

  ➢ The model has access only to the input features during testing, not the labels.

  ➢ Test samples provide an unbiased evaluation of the model's ability to generalize to new, unseen data.

HAMK
Häme University
of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin
osarahoittama

How do we use training data now?

- Take new hobbit as test sample.

- We do not use training samples for testing, because the model has seen these samples, and it does not give us idea how well it learnt.

- We take the test sample and estimate its height (y).
  - We get error of test sample!
  - y_true – y_estimate = y_true – (a*x_test+b)

Ground truth

N samples

x1 → y1

x2 → y2

x3 → y3

....

xN → yN

How we call this
kind of system?

Stick to high
school math

N samples

x1 → y1

x2 → y2

x3 → y3

….

xN → yN

Turn into **system of linear equations**

N samples

x1 → y1

x2 → y2

x3 → y3

….

xN → yN

y1 = a*x1 + b

y2 = a*x2 + b

y3 = a*x3 + b

….

yN = a*xN + b

N equations

Häme University of Applied Sciences

Euroopan unionin osarahoittama

N samples

x1 → y1

x2 → y2

x3 → y3

....

xN → yN

y1 = a*x1 + b

y2 = a*x2 + b

y3 = a*x3 + b

....

yN = a*xN + b

Optimal a and b?

Optimal in which sense

How to calculate and what to optimize?

Minimize error

- **Error** in ML is also called **loss**
- **Minimize error** for all our points (residual value between ground truth and our estimate)
- y_true_i – y_estimate_i,   (i samples)
- → get sum of all errors:
- $\sum_{i}$ (y_true_i – y_estimate_i)

Minimize error

- Error in ML also called loss
- Minimize error for all our points (residual value between ground truth and our estimate)
- y_true_i – y_estimate_i,   (i samples)
- → get sum of all errors:
- $\sum_{i}$ (y_true_i – y_estimate_i)

What is the problem of sum?

Some positive, some negative compensate each other

May result in zero error in the end. Perfect fit!?

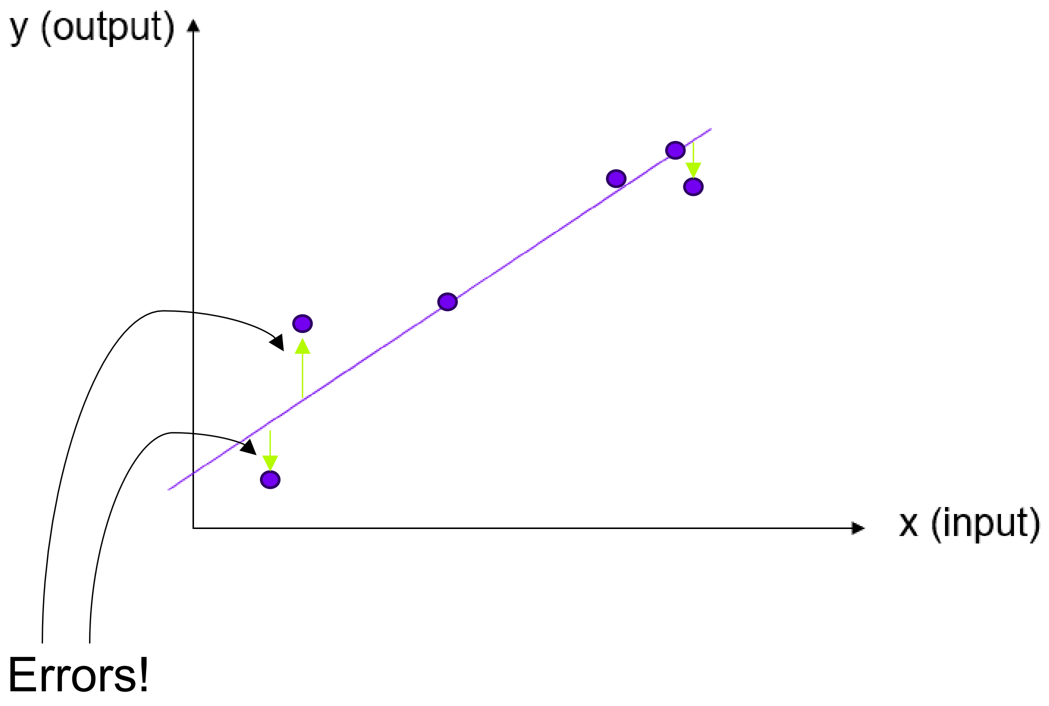
HAMK
Häme University
of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin
osarahoittama

Linear regression

- **$y = ax + b$**
- Some errors below the line, some are above



HAMK
Häme University
of Applied Sciences

Euroopan unionin
osarahoittama

Minimizing error

- $\sum_i$ (y_true_i – y_estimate_i)

Take absolute value?

Minimizing error

- $\sum_i$ (y_true_i – y_estimate_i)

Take absolute value?

Engineers don't like absolute values

|0| → undefined gradient

Minimizing error

- $\sum_{i} (y\_true\_i - y\_estimate\_i)^2$

Take square ^2

Always positive

Minimizing error

- $\sum_i$ (y_true_i – y_estimate_i)^2

Take square ^2

Problems?

Minimizing error

- $\sum_{i} (y\_true\_i - y\_estimate\_i)^2$

Take square ^2

Problems?

Error of 10 points vs 100 points

Minimizing error

- $\sum_i (y\_true\_i - y\_estimate\_i)^2$

Take square ^2

Problems?

Error of 10 points vs 100 points

Error of 100 points bigger as the number of values is more

How to avoid this problem?

HAMK Häme University of Applied Sciences

REGIONAL UNIVERSITY NETWORK EUROPEAN UNIVERSITY

Euroopan unionin osarahoittama

Mean squared error

$$\mathrm{MSE} = \overset{\text{Mean}}{\boxed{\frac{1}{n}\sum_{i=1}^{n}}}\Big(\overset{\text{Error}}{\boxed{\underset{\text{Ground truth}}{Y_i} - \underset{\text{Estimate}}{\hat{Y}_i}}}\Big)^{\overset{\text{Squared}}{\boxed{2}}}$$

Take mean value!

Number of points doesn't matter anymore

HAMK
Häme University
of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin
osarahoittama

Two rules of thumbs for engineers when you don’t know what to do

- Always fit the line.
- If need to measure error, measure MSE.

$$MSE = \frac{\sum_{i=1}^{N} \left(y_i - (ax_i + b)\right)^2}{N}$$

Find a and b so that error is minimized to the given training data (<x_i, y_i>,...)

- The problem is defined now.

Find a and b so that error is minimized to the given training data (<x_i, y_i>,...)

- The problem is defined now.
- How to solve the problem?

HAMK
Häme University of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin osarahoittama

Find a and b so that error is minimized to the given training data (<x_i, y_i>,...)

- **Solution 1** (for lazy): **Brute force**
  - Test random values or all values for a and b
  - Select the best in terms of MSE
  - Test thousands values in one second

for a = -100 by 0.1 to 100
for b = -100 by 0.1 to 100

Find a and b so that error is minimized to the given training data (<x_i, y_i>,...)

- **Solution 1** (for lazy): **Brute force**
  - Test random values or all values for a and b
  - Select the best in terms of MSE
  - Test thousands values in one second

for a = -100 by 0.1 to 100
for b = -100 by 0.1 to 100

Slow and in some cases this might take ages ☹

If use Brute force in all cases, we would need 1000 power plants more to power all computers

Something more elegant?

HAMK Häme University of Applied Sciences

REGIONAL UNIVERSITY NETWORK EUROPEAN UNIVERSITY

Euroopan unionin osarahoittama

Find a and b so that error is minimized to the data (<x_i, y_i>,...)

- **Solution 2:**

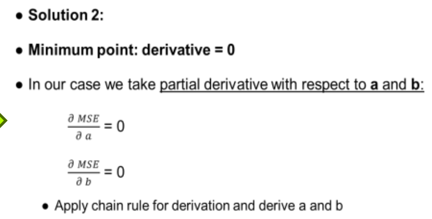
- **Minimum point: derivative = 0**

- In our case we take <u>partial derivative with respect to **a** and **b**:</u>

$$\frac{\partial\, MSE}{\partial\, a} = 0$$

$$\frac{\partial\, MSE}{\partial\, b} = 0$$

  - Apply chain rule for derivation and derive a and b



HAMK
Häme University
of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin osarahoittama

Linear regression formulas

$$a = \frac{[(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)]}{[n(\Sigma x^2) - (\Sigma x)^2]}$$

$$b = \frac{[n(\Sigma xy) - (\Sigma x)(\Sigma y)]}{[n(\Sigma x^2) - (\Sigma x)^2]}$$

Let’s test them in Python

Fit line task

- Machine Learning models often rely on fitting data to patterns. One of the most fundamental models is linear regression, which **fits a straight line to data points.**
- **Exercise:** interactively collect training data by clicking on a plot and then fit a linear model to it. **Practice:** ML, event-driven programming, numpy-based math.
- **Questions**:
  - ❑ What would happen if all the input points had the same x-coordinate?
  - ❑ How is this related to training a machine learning model?

Linear regression

- Questions:
  - ❑ **What would happen if all the input points had the same x-coordinate?**
  - Every x_i = c, for some constant c.
  - Therefore x_estimated = c, and x_i – x_estimated = 0 for all i → **slope a is undefined** (singular case):

$$\sum(x_i - \bar{x}) = 0, \quad \sum(x_i - \bar{x})^2 = 0$$

$$a = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{0}{0}$$

Linear regression

- Questions:
  - ❑ What would happen if all the input points had the same x-coordinate?

$$a = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{0}{0}$$

  - ❑ **How is this related to training a machine learning model?**
  - The data lacks variance in feature – it's non-informative. We cannot learn a meaningful model: **there's no information in x to predict y.**
  - Linear models fail to train.

Extend linear regression to linear classification.

HAMK
Häme University
of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

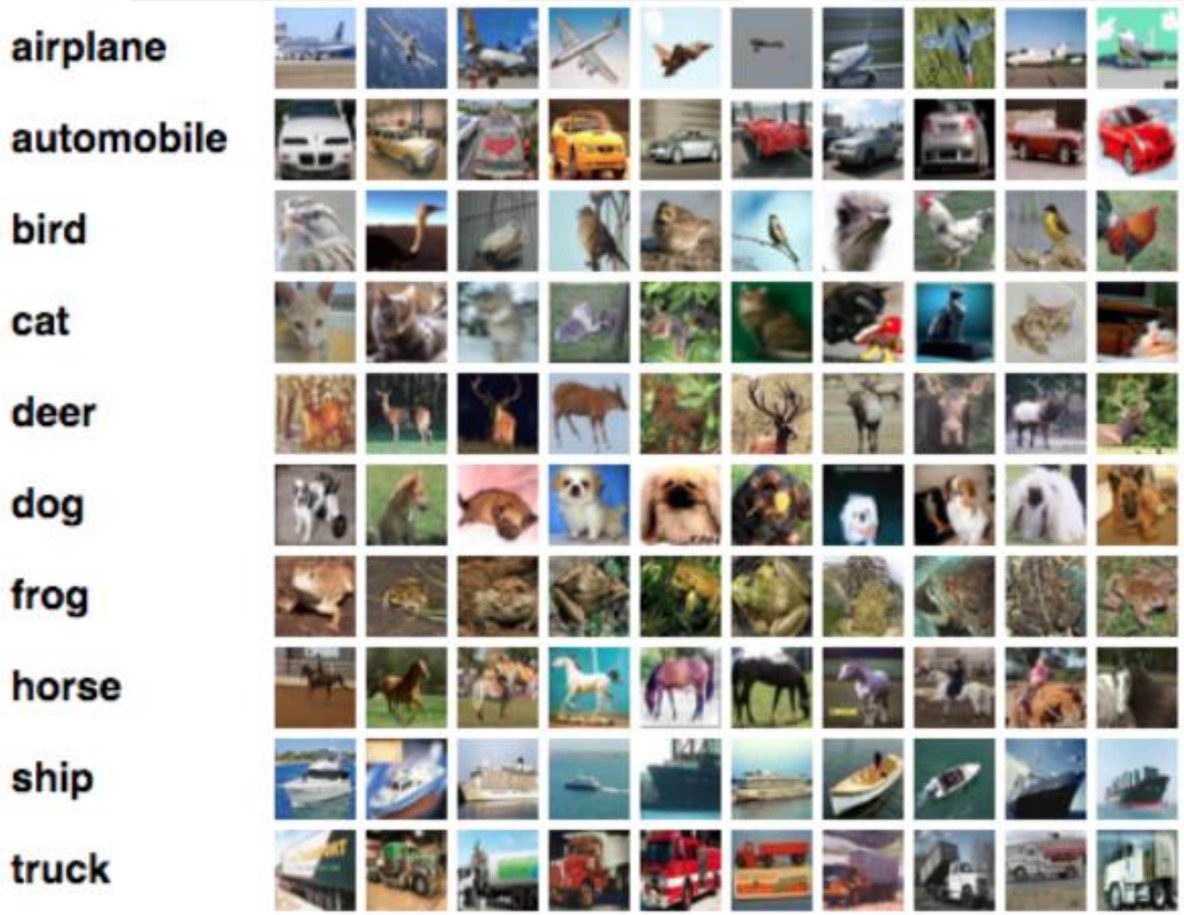
Euroopan unionin
osarahoittama

Linear classification and classification problem

- The problem of Machine Learning – how it all started.
- Problem of assigning input $\bar{x}$ to one of the pre-defined classes y={1, 2, 3} (“car”, “airplane”, “bird”).
- Here $\bar{x}$ (with bar) denotes not a single value, but some kind of input that we have, it can be anything.
- $\bar{x}$: observation (sample), can be value, vector, image, audio sample (based on the image, video, or sound model defines the class number).

Euroopan unionin osarahoittama

Linear classification and classification problem



- Similar to regression.
- The solution is learnt from training data.

x1 → y1 (known class, ground truth)
x2 → y2
x3 → y3
….

HAMK
Häme University
of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin
osarahoittama

Example from Middle-Earth: Hobbit vs elf

- Orcs love to eat hobbits, so they want to make a trap to <u>catch hobbits</u>, but <u>not elves</u>, as orcs are afraid of elves.
- One clever orc who uses Machine Learning wants to **create a classifier**.
- Hobbits vs elves:

How to distinguish between these two classes?

Example from Middle-Earth: Hobbit vs elf

- Orcs love to eat hobbits, so they want to make a trap to catch hobbits, but not elves, as orcs are afraid of elves.
- One clever orc who knows Machine Learning wants to create a classifier.
- Hobbits vs elves:

Hobbits are short and elves are tall

Hobbit vs elf

- Measurement to tell the height of someone walking -> this can be our classifier.
- x: height

Hobbit vs elf

- Measurement to tell the height of someone walking -> this can be our classifier.
- x: height

How we measure the height in this case?

Hobbit vs elf

- Measurement to tell the height of someone walking -> this can be our classifier.
- x: height

How we measure the height in this case?

Light emitting bars that are blocked with the height

HAMK
Häme University
of Applied Sciences

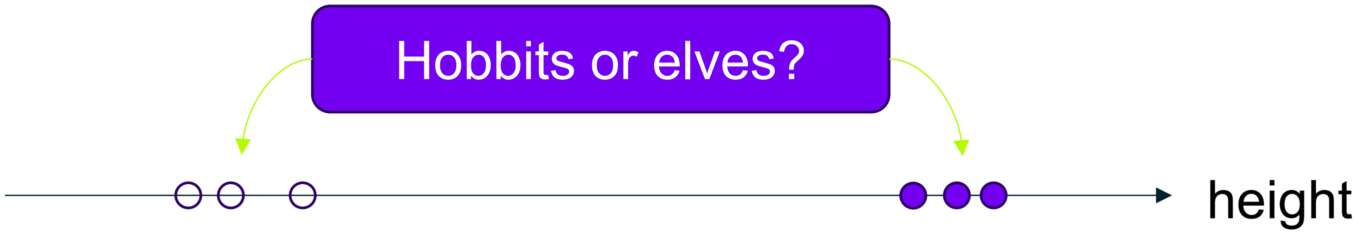
REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin
osarahoittama

Hobbit vs elf

- Measurement to tell the height of someone walking → this can be our classifier.
- x: height
- y: {hobbit, elf}

Light emitting bars that are blocked with the height



HAMK Häme University of Applied Sciences

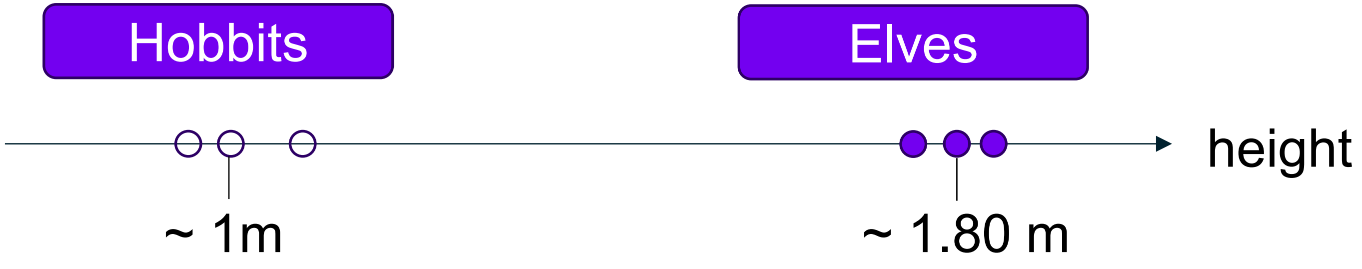
REGIONAL UNIVERSITY NETWORK EUROPEAN UNIVERSITY

Euroopan unionin osarahoittama

Hobbit vs elf

- Measurement to tell the height of someone walking -> this can be our classifier.
- x: height
- y: {hobbit, elf}

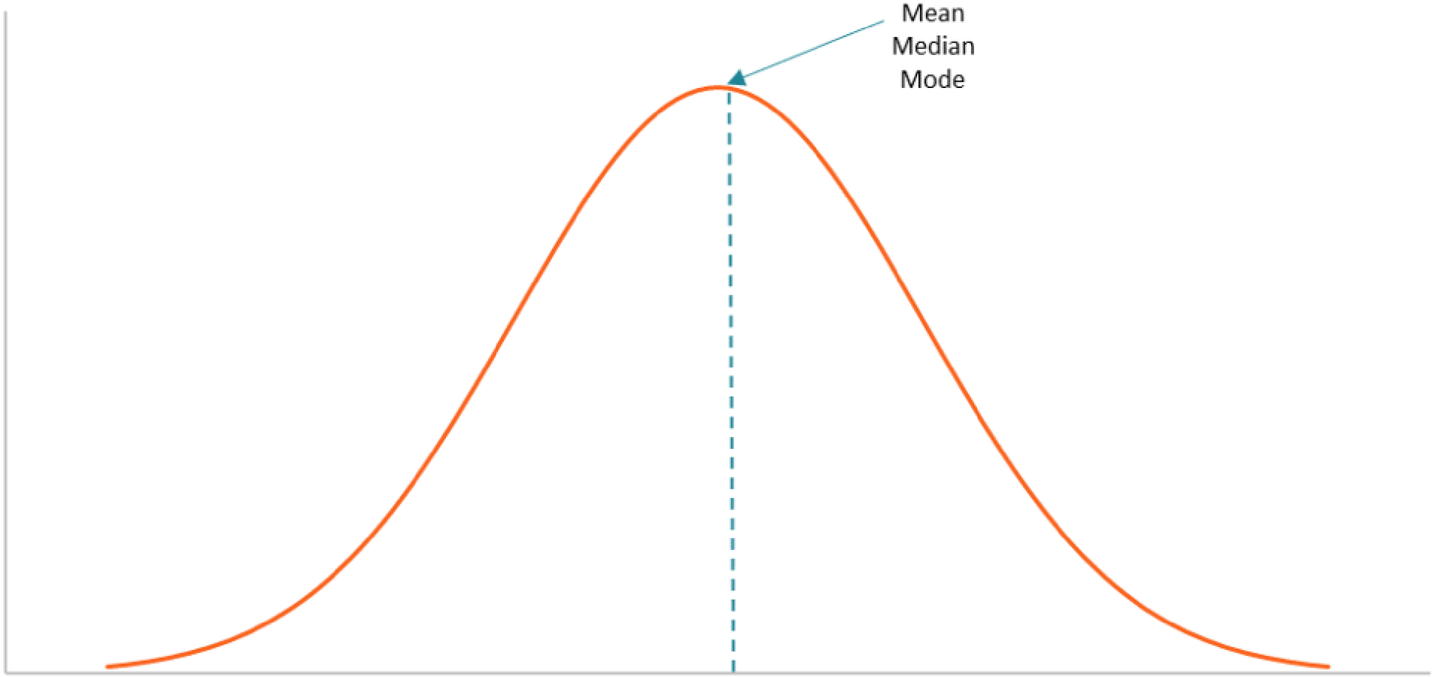
Light emitting bars that are blocked with the height



How height of hobbits and elves will be distributed?

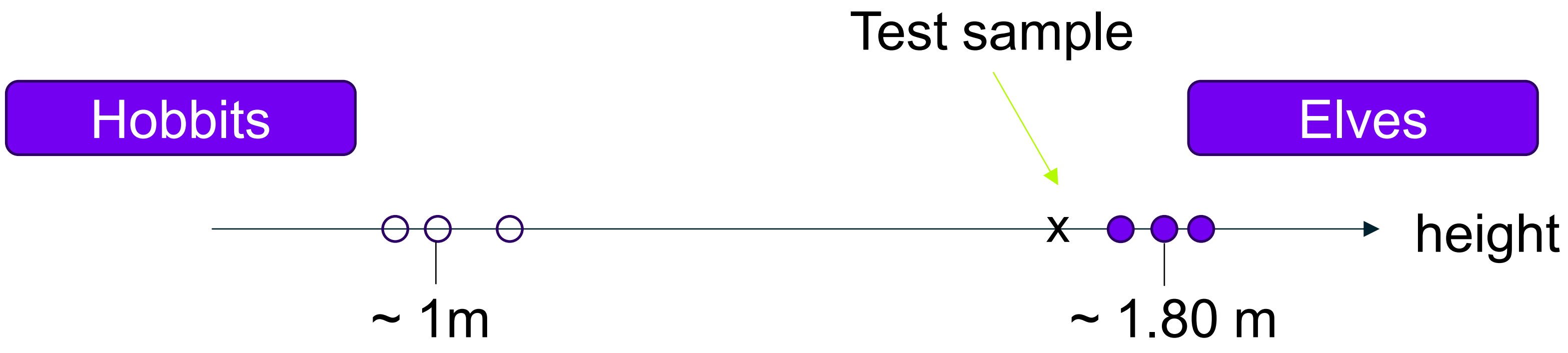
How height of hobbits and elves will be distributed?

- Quite many things in the nature are normally distributed:
  - Bell curve



New sample: how to classify it?

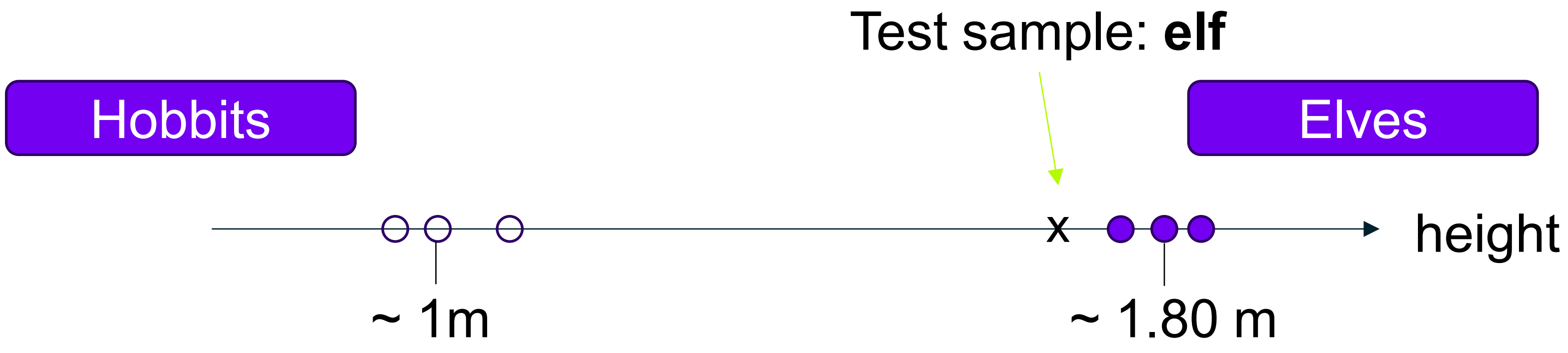
The problem of classification:



New sample: how to classify it?

The problem of classification:

This is probably elf, but what is the intuition behind this classification?

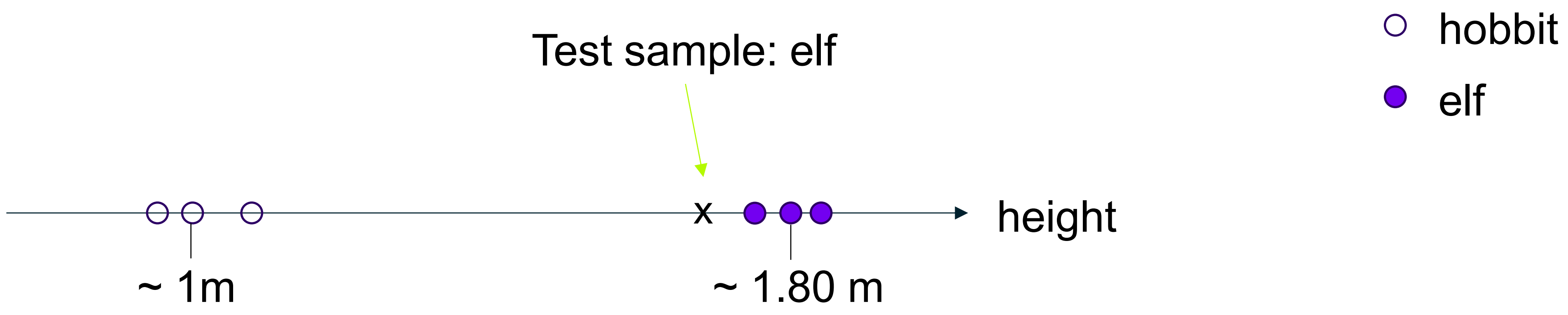


HAMK
Häme University
of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin osarahoittama

New sample: how to classify it?



! Nearest Neighbor Classifier

Nearest Neighbor Classifier

- **Learning**: store all training sample.
- **Inference** (making the decision for some sample): measure distance to all training samples and return class of the shortest distance.

HAMK
Häme University
of Applied Sciences

Euroopan unionin
osarahoittama

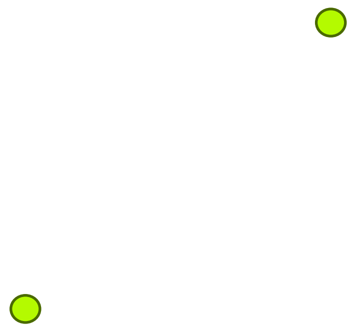
Distance

- What is distance?
- What is typical way to measure distance?

Euroopan unionin
osarahoittama

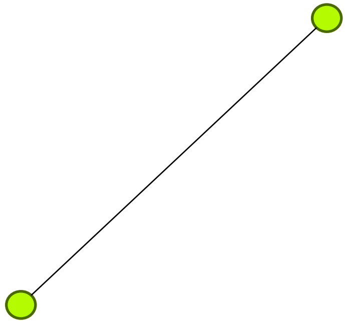
Distance

- Measure distance between two points



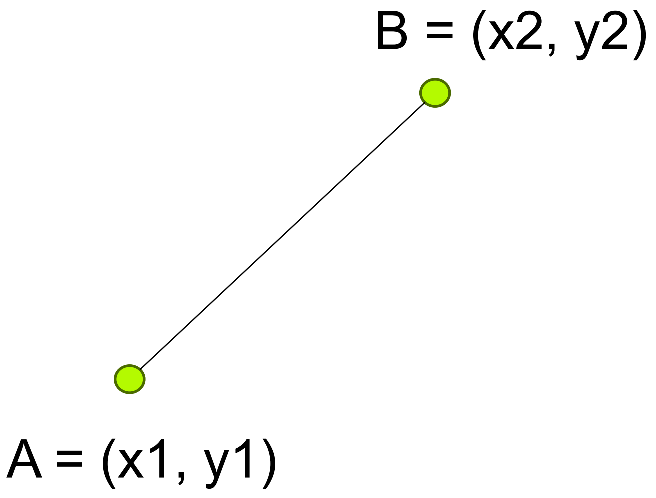
Distance

- Measure distance between two points:



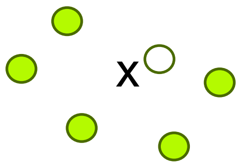
Distance

- Measure distance between two points.
- **Euclidean distance** or L2 distance:
  - Distance(A, B) = $\sqrt{(x1 - x2)^2 + (y1 - y2)^2}$



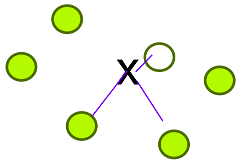
Nearest neighbor

- Take a sample x
    - Close to ○
    - Surrounded by ●
- How to make our classifier robust?



Nearest neighbor

- Take a sample x
  - Close to
  - Surrounded by
- How to make our classifier **robust**?
- Instead of 1-NN -> 3-NN and **majority** vote.



Pros:
Easy and intuitive

Cons:
Slow (imagine 8 million samples)

HAMK
Häme University
of Applied Sciences

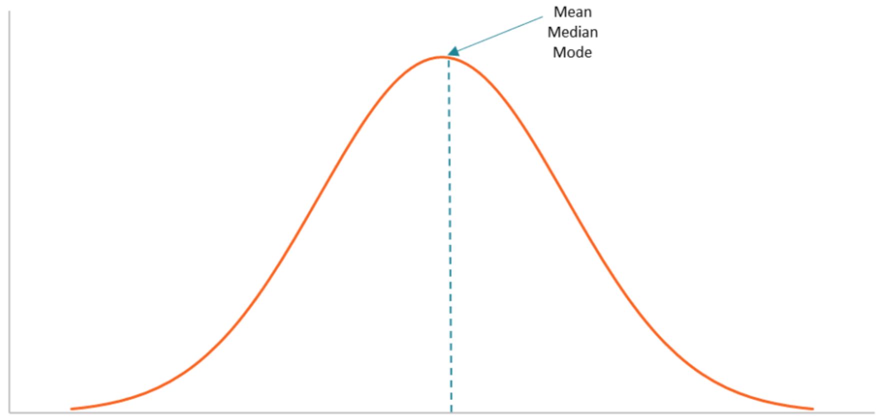
REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin osarahoittama

Back to linear classification

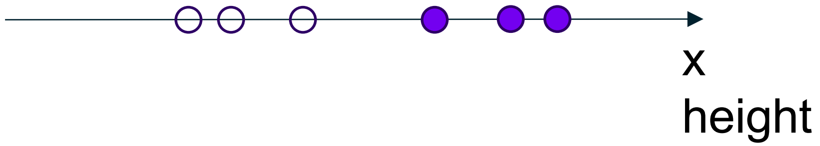
- Engineering (or Newtonian) approach:
- Represent the problem as points in the space and use method of linear regression (fitting a line) to solve the problem.
- Rule #1: always fit the line.
- Rule #2: MSE.
- Rule #3: if you do not know anything about distribution, assume that distribution is Gaussian.

$$MSE = \frac{\sum_{i=1}^{N} \left(y_i - (ax_i + b)\right)^2}{N}$$



Fitting a line

- Measurements:

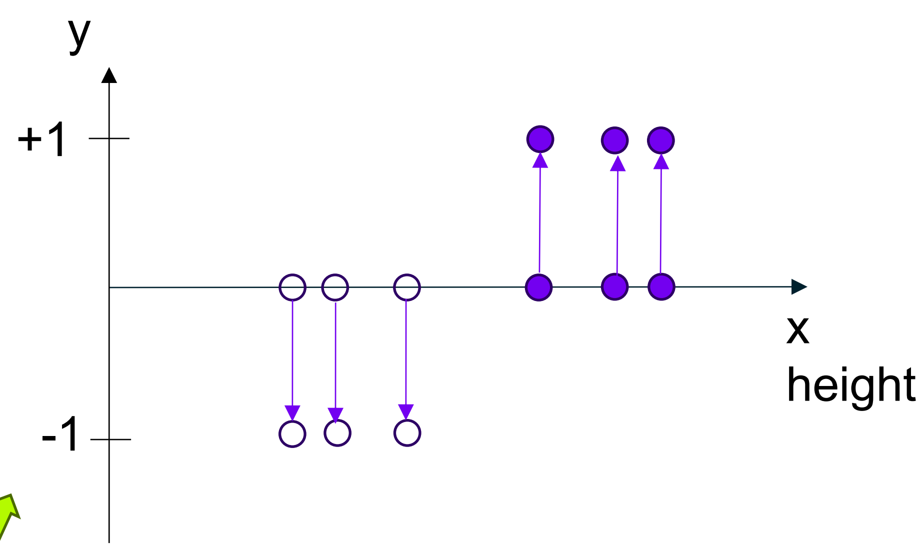


Fitting a line

- Measurements:
  - They are on the same line x
  - If we fit the line, it will go through all those points.

Fitting a line

- Measurements:
  - They are on the same line x.
  - If we fit the line, it will go through all those points.
  - y: represents class label.



We invent numbers +1 and -1 to represent two different classes and move all points

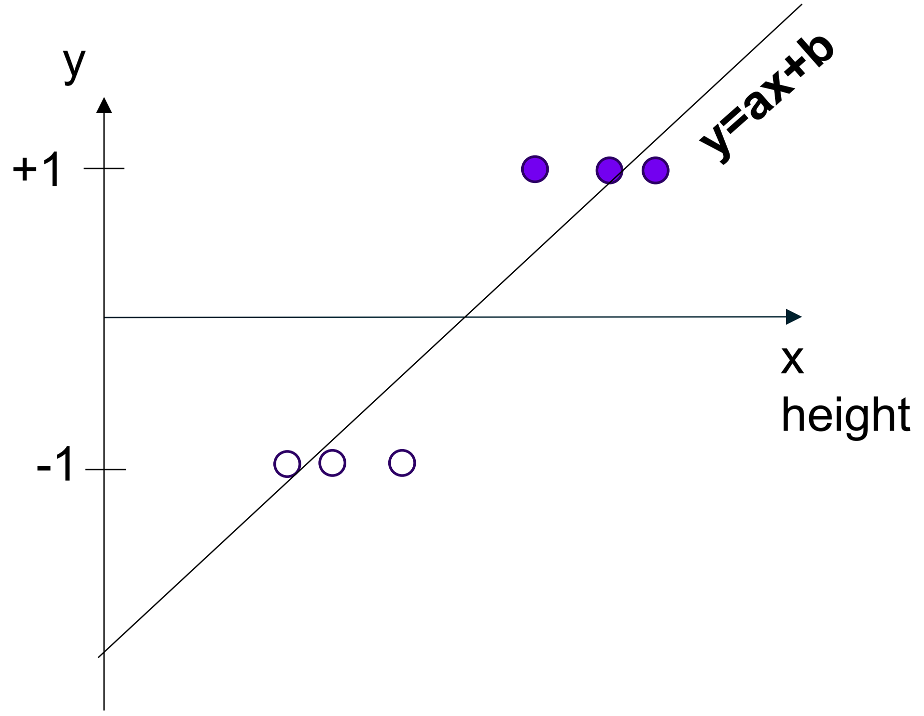
HAMK
Häme University of Applied Sciences

RUN Regional University Network
European University

Euroopan unionin osarahoittama

Fitting a line

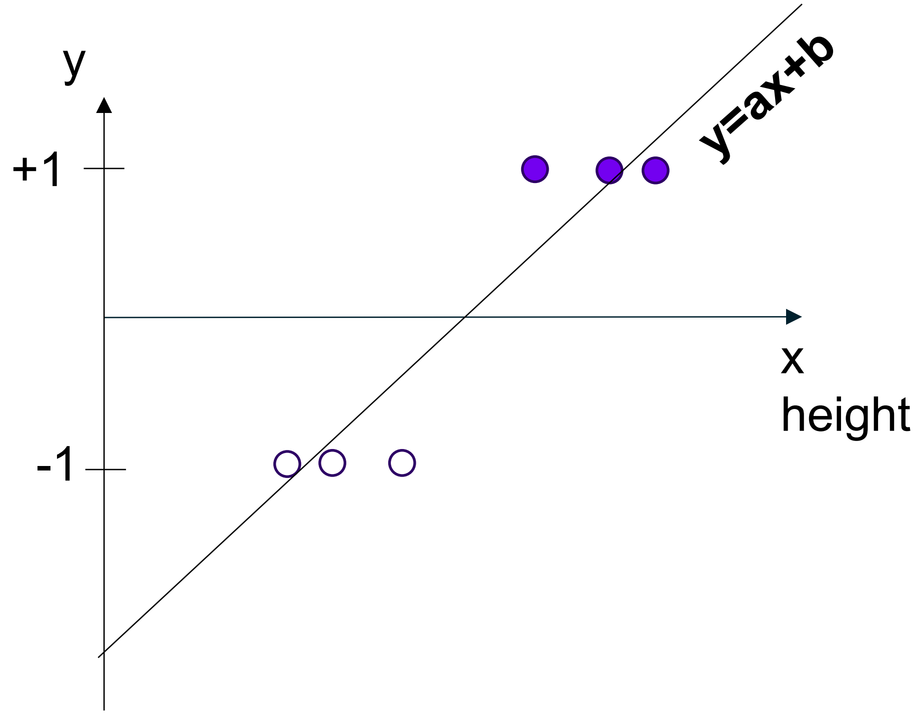
- Measurements:
  - They are on the same line x.
  - If we fit the line, it will go through all those points.
  - y: represents class label.



Now we can fit the line that goes through all these points

Fitting a line

- Measurements:
  - They are on the same line x.
  - If we fit the line, it will go through all those points.
  - y: represents class label.

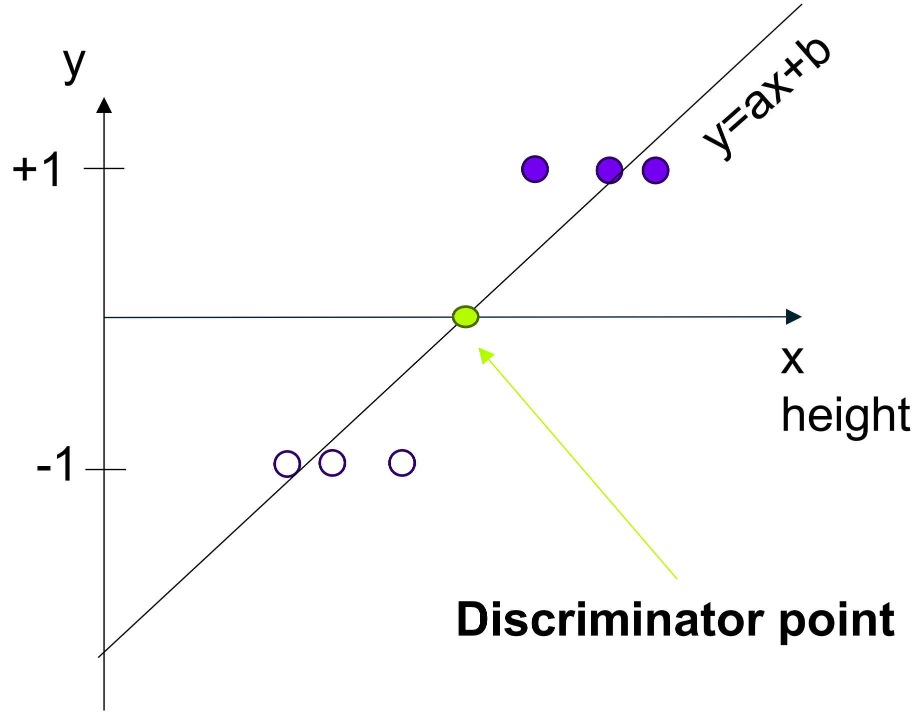


Now we can fit the line that goes through all these points

How do we use it as a classifier?

Fitting a line

- Measurements:
  - They are on the same line x.
  - If we fit the line, it will go through all those points.
  - y: represents class label.



Now we can fit the line that goes through all these points

How do we use it as a classifier?

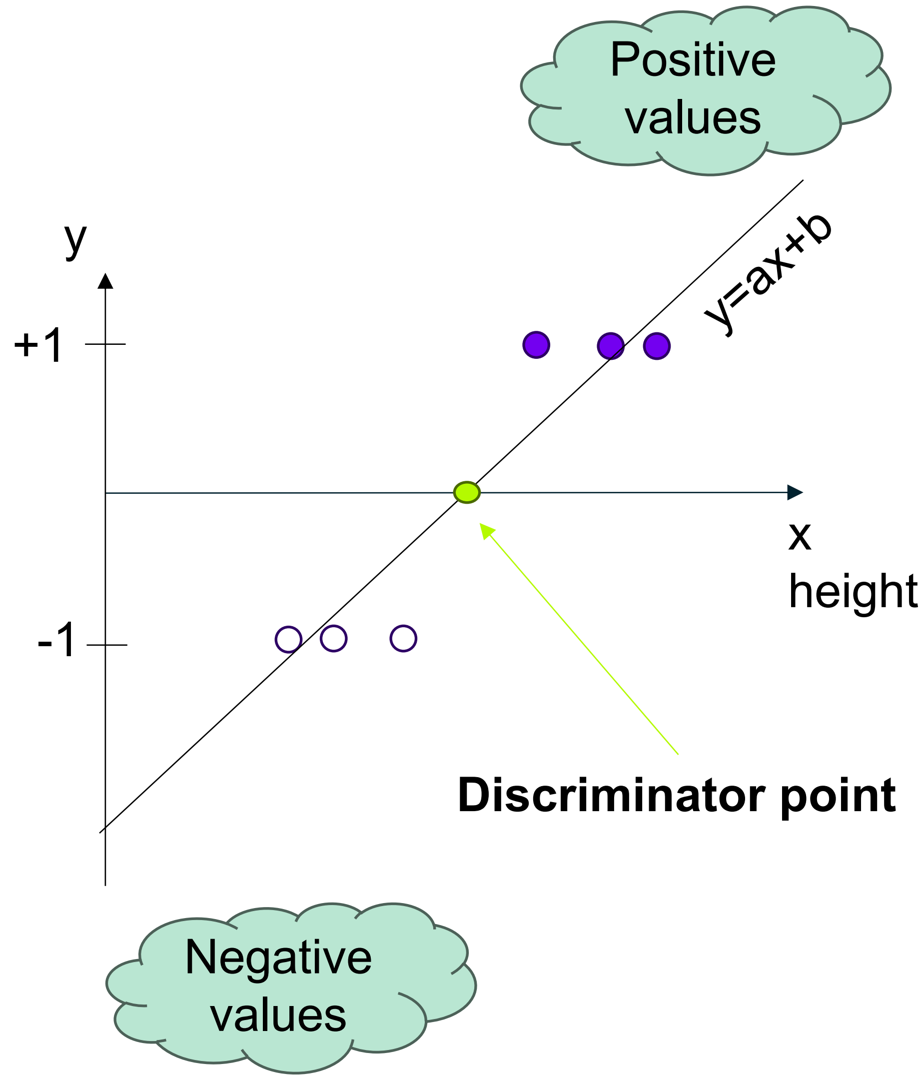
HAMK
Häme University of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin osarahoittama

Fitting a line

- Measurements:
    - They are on the same line x.
    - If we fit the line, it will go through all those points.
    - y: represents class label



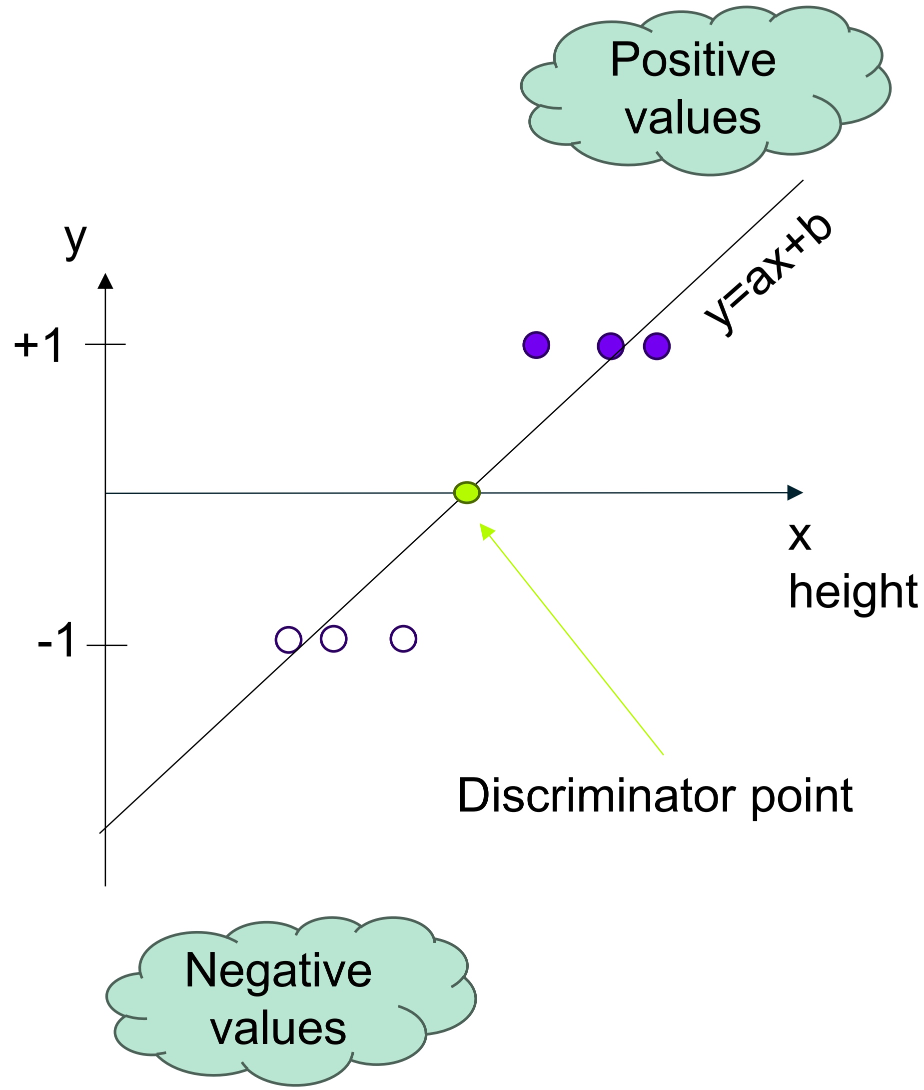
Classifier rule:
Hobbit: if ax+b < 0
Elf: if ax+b > 0

HAMK
Häme University of Applied Sciences

Euroopan unionin osarahoittama

Fitting a line

- Measurements:
  - They are on the same line x.
  - If we fit the line, it will go through all those points.
  - y: represents class label.



<u>Classifier rule:</u>
Hobbit: if ax+b < 0
Elf: if ax+b > 0

Problem solved, but where we put = ?
When we are exactly at the discriminator point?

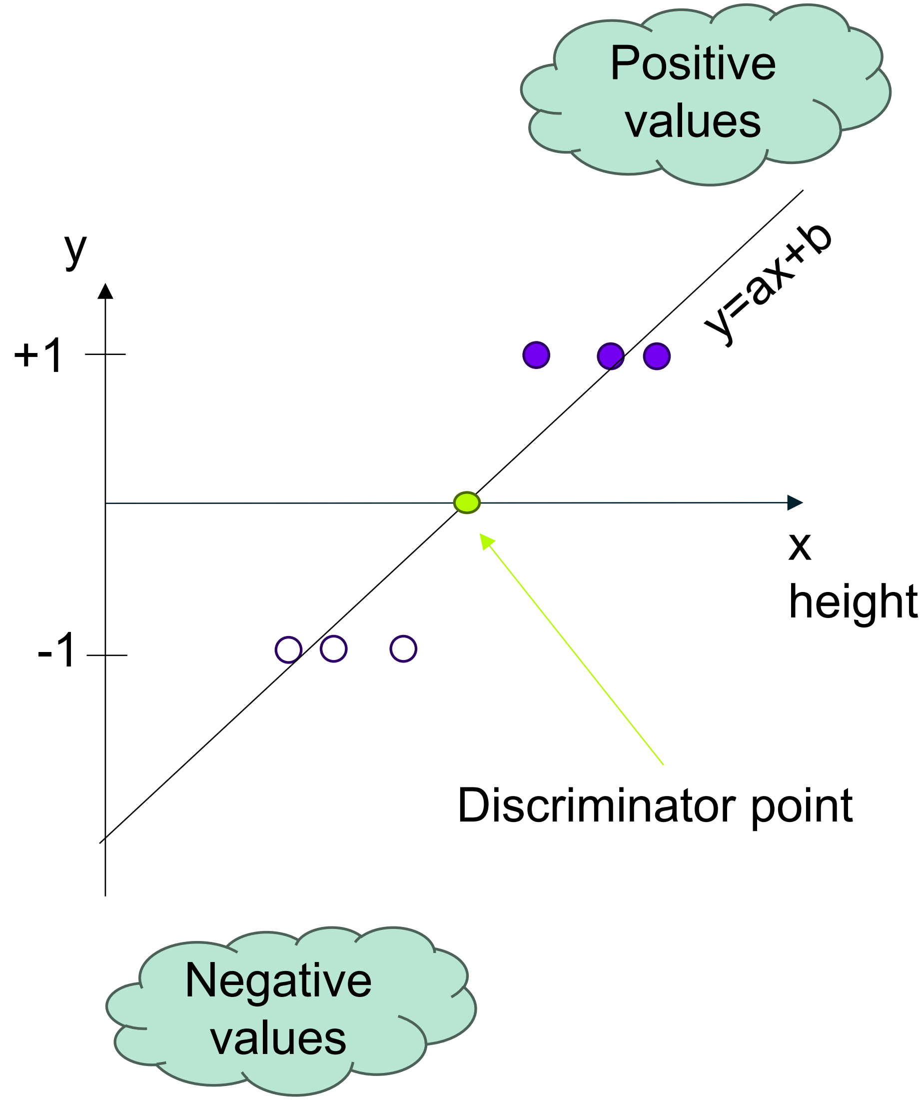
Toss a coin?

Fitting a line

<u>Classifier rule:</u>
Hobbit: if ax+b < 0
Elf: if ax+b > 0

Problem solved, but where we put = ?
When we are exactly at the the discriminator point?

Pick the class with more points



HAMK Häme University of Applied Sciences

REGIONAL UNIVERSITY NETWORK EUROPEAN UNIVERSITY

Euroopan unionin osarahoittama

Discriminator point in Middle-Earth

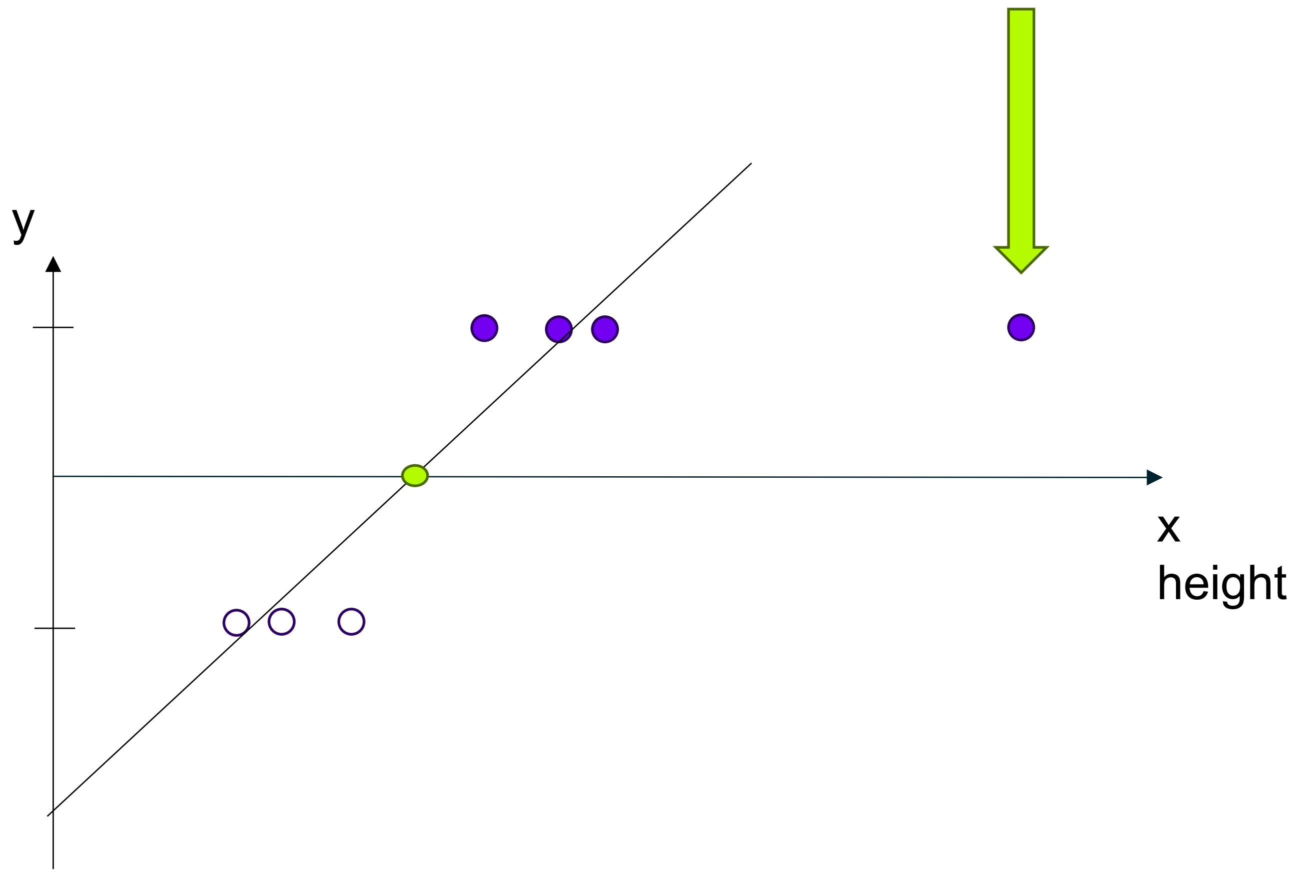
- Out of 10 creatures that pass by, 9 are elves and only 1 is hobbit.
- So, we assume it is elf.
- Or, put equal mark to elves, to avoid danger and minimize the risk to catch elf.
- Problem solved.
- **Any troubles?**

HAMK
Häme University
of Applied Sciences

Euroopan unionin
osarahoittama

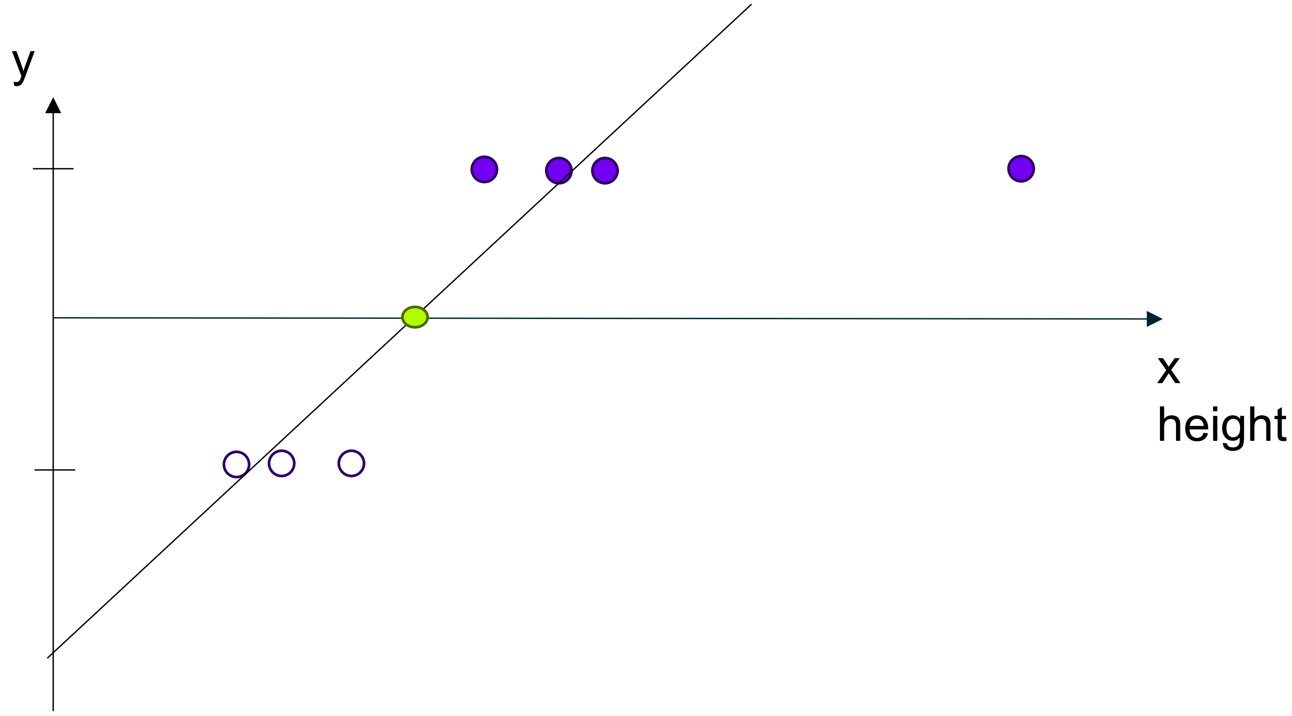
The trouble

- Let’s assume one giant elf.
- With the current classifier rule, how the giant elf is classified?



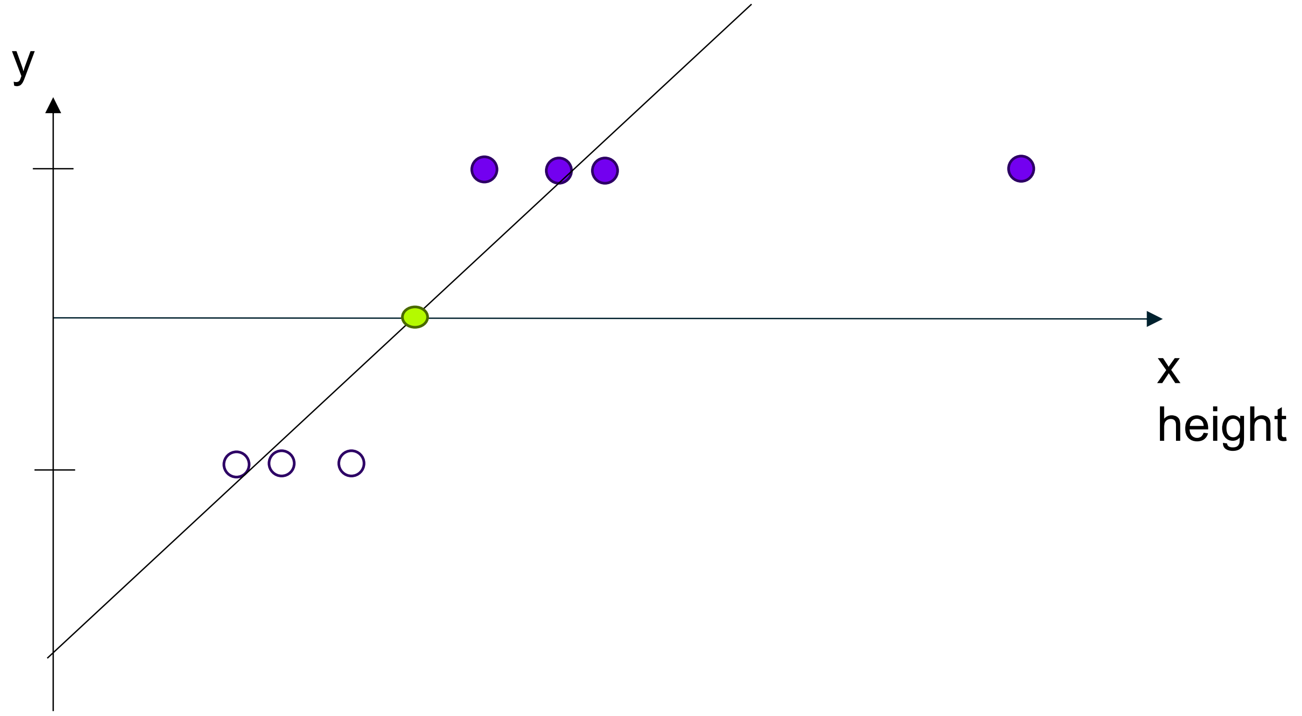
Trouble

- Let's assume one giant elf.
- With the current classifier rule, the giant elf is classified **correctly**.



Trouble

- Let’s assume one giant elf.
- With the current classifier rule, the giant elf is classified correctly.
- But what happens if we **add giant elf to training set** and refit the line?



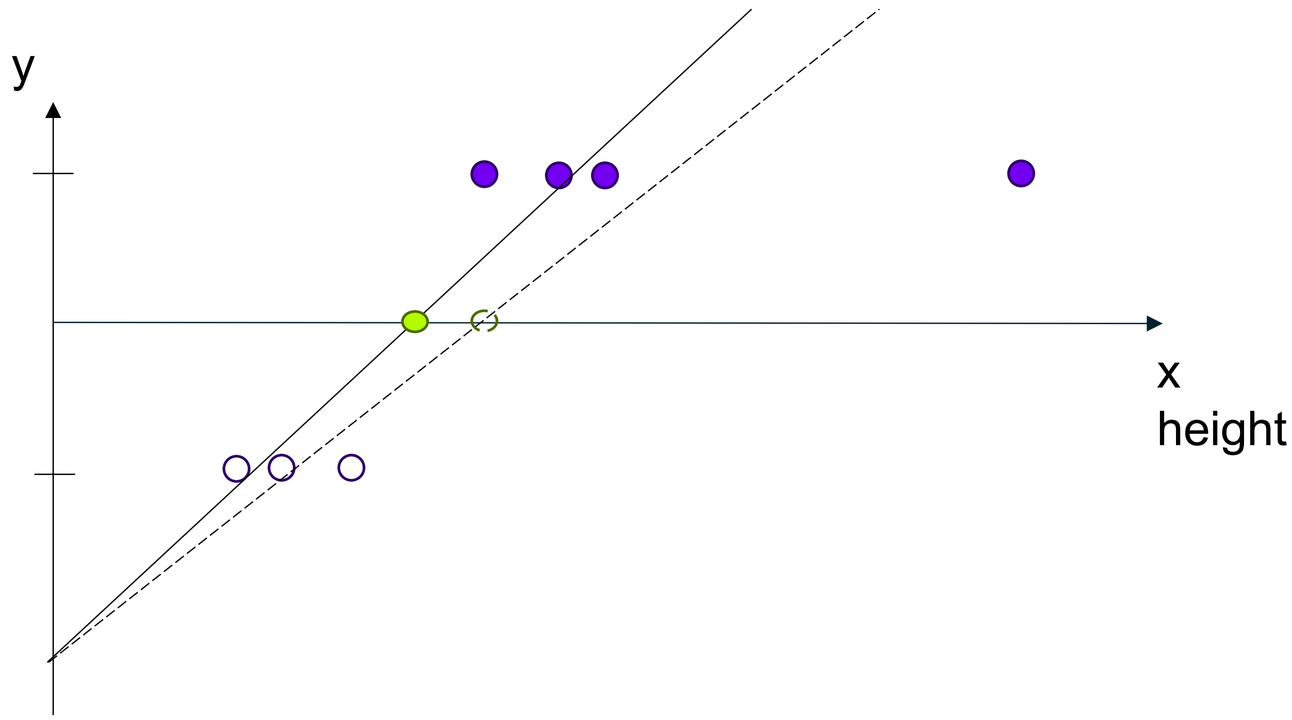
HAMK Häme University of Applied Sciences

REGIONAL UNIVERSITY NETWORK EUROPEAN UNIVERSITY

Euroopan unionin osarahoittama

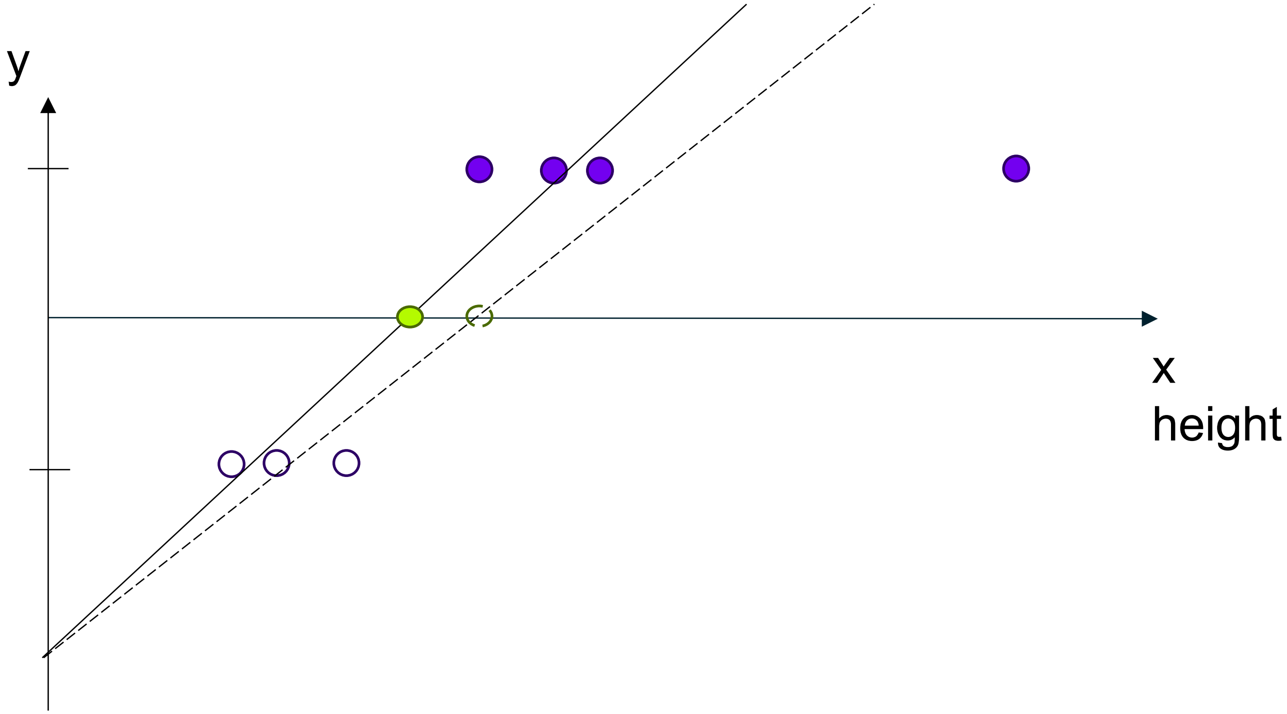
Trouble

- The **new line** bends to the right.
- Some of elves might slip to the left and we misclassify them.



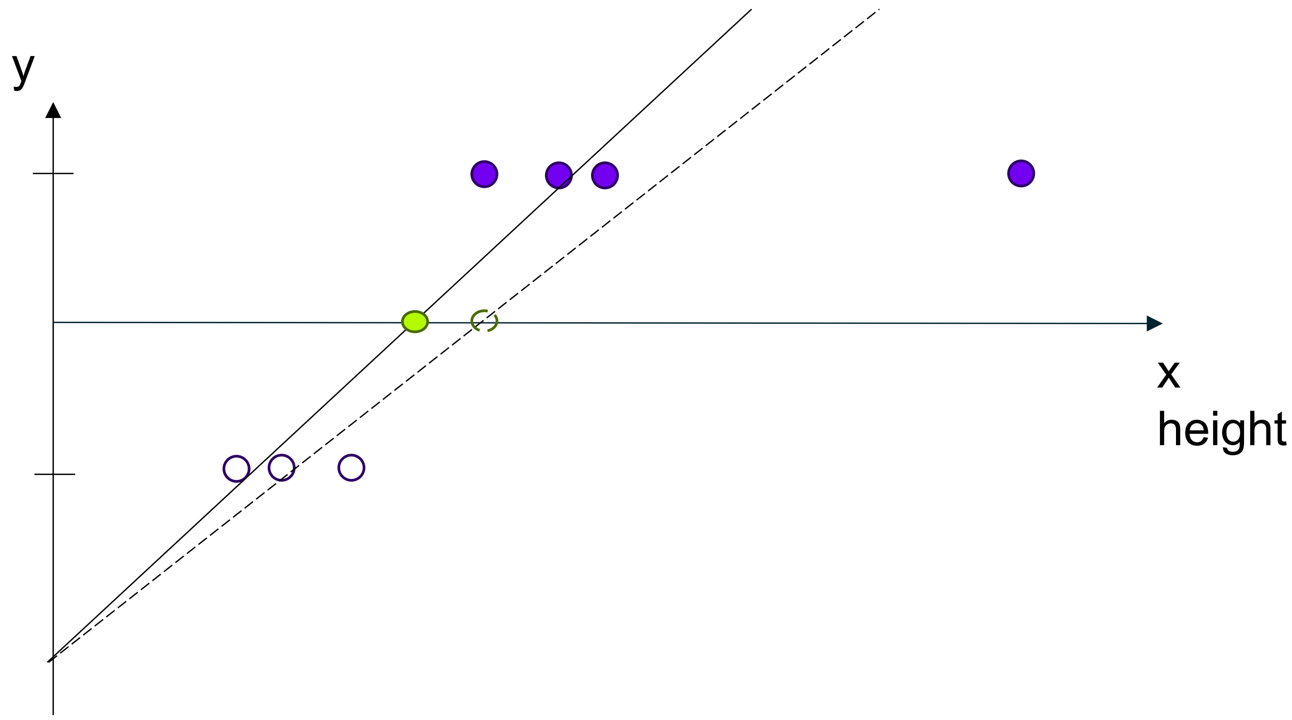
Trouble

- The new line bends to the right.
- Some of elves might slip to the left and we misclassify them.
- **Fundamental problem in our classifier: how to define error for classification?**



Trouble

- How to define error for classification?
- The output class is either wrong or correct.



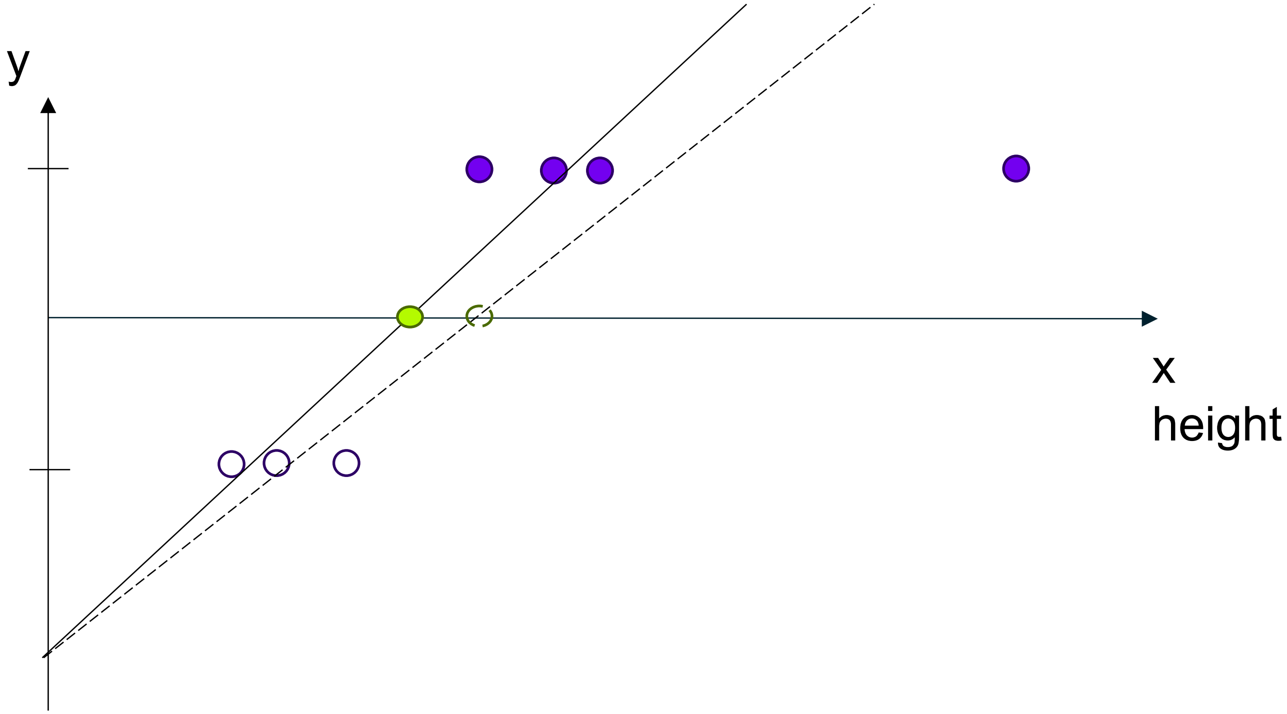
HAMK Häme University of Applied Sciences

REGIONAL UNIVERSITY NETWORK EUROPEAN UNIVERSITY

Euroopan unionin osarahoittama

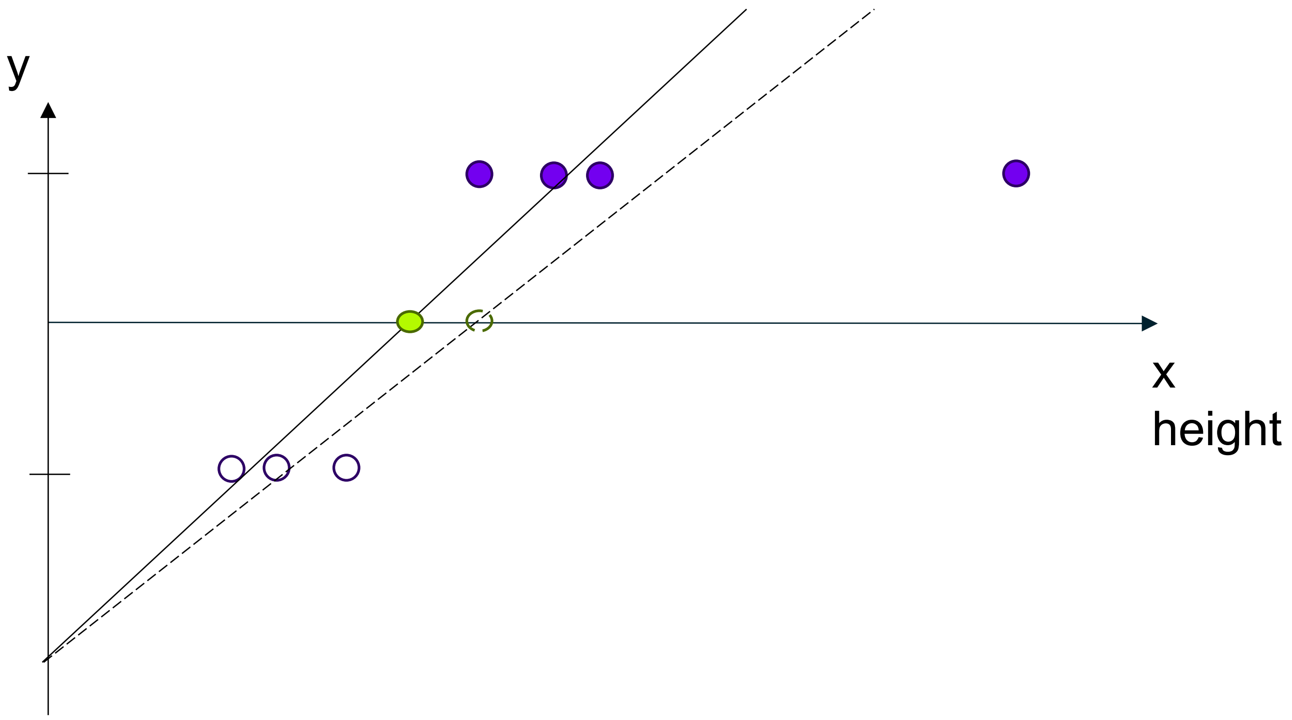
Trouble

- How to define error for classification?
- The output class is either wrong or correct.
- Let's assume correct label:
  - MSE = ?



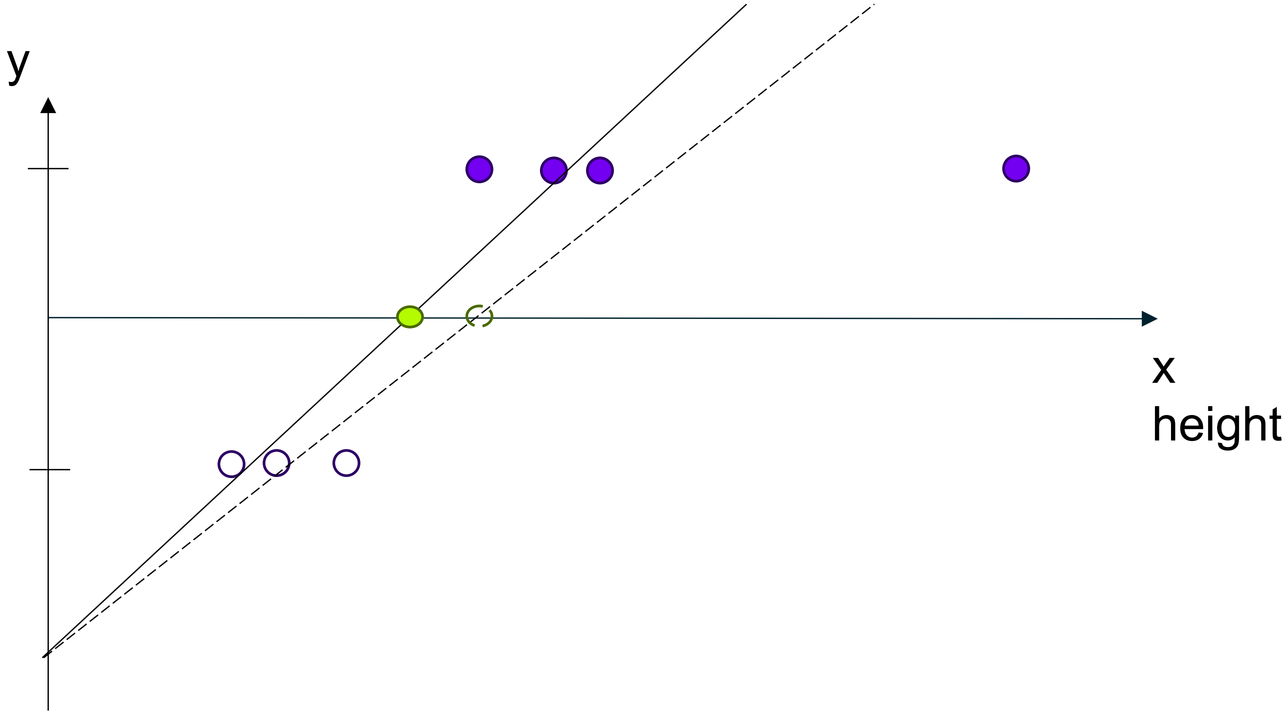
Trouble

- How to define error for classification?
- The output class is either wrong or correct.
- Let’s assume correct label:
  - MSE = 0
- Wrong label:
  - MSE = ?



Trouble

- How to define error for classification?
- The output class is either wrong or correct.
- Let's assume correct label:
  - MSE = 0
- Wrong label:
  - MSE = 1(intuitively)

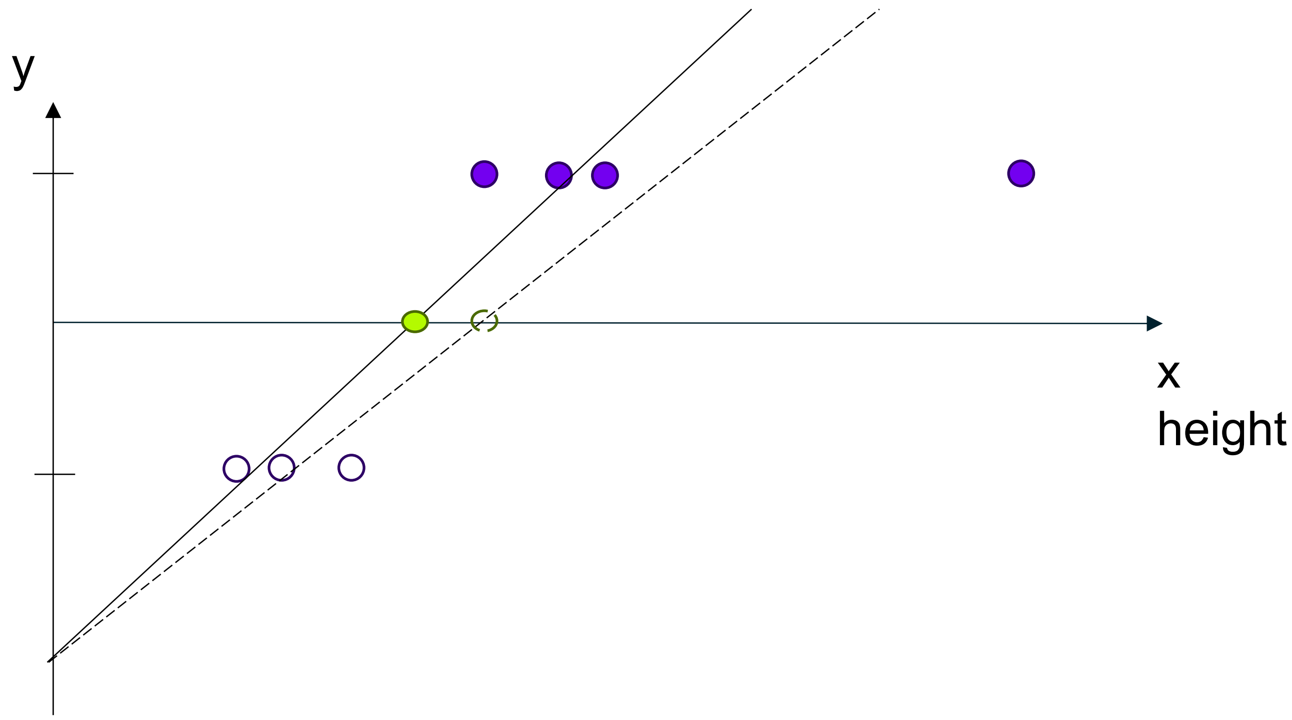


100 samples and MSE = 20

Trouble

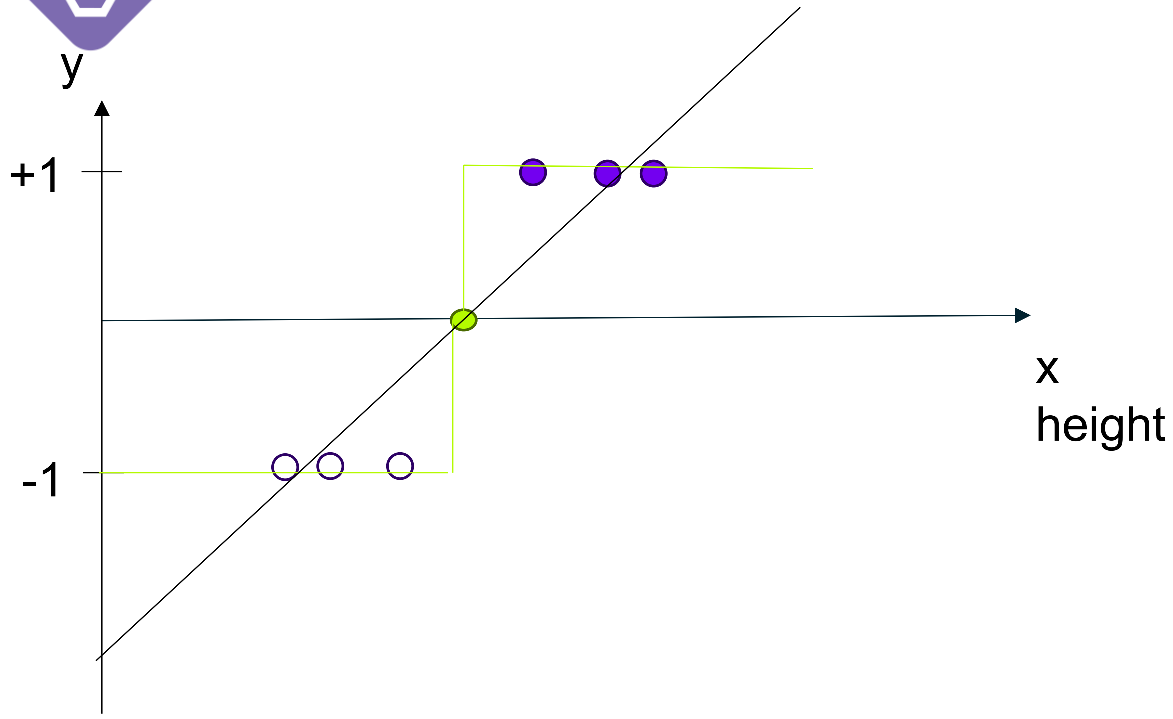
- How to define error for classification?
- The output class is either wrong or correct.
- Let's assume correct label:
  - MSE = 0
- Wrong label:
  - MSE = 1

100 samples and MSE = 20, we know that 20 samples were misclassified



Solving the trouble

Discretize the problem

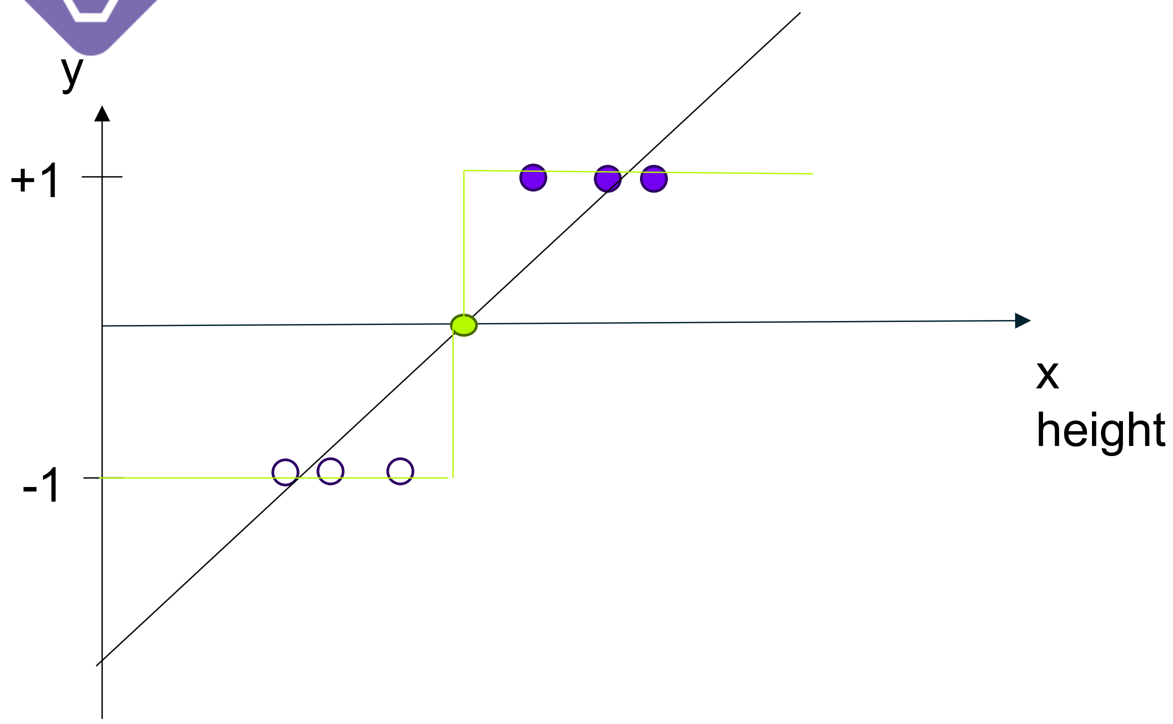


HAMK
Häme University
of Applied Sciences

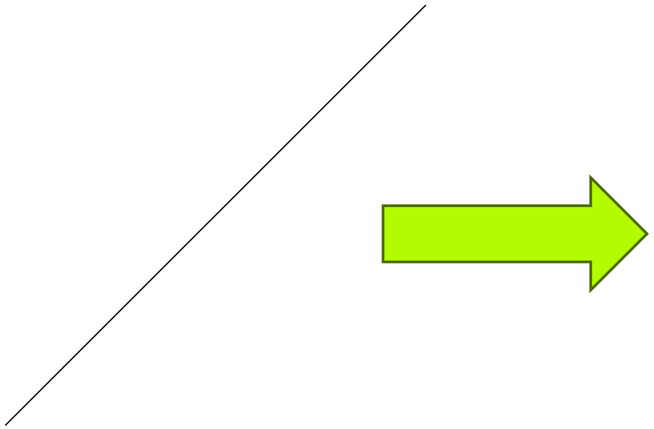
REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin
osarahoittama

Solving the trouble



Discretize the problem

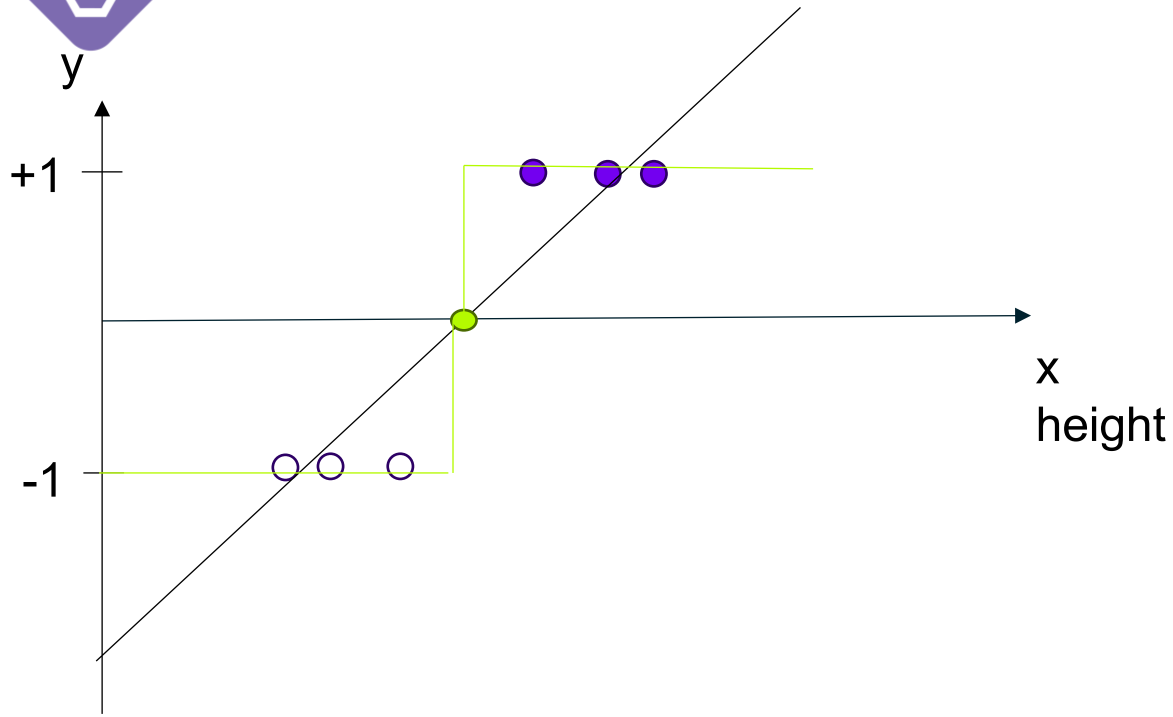


HAMK
Häme University
of Applied Sciences

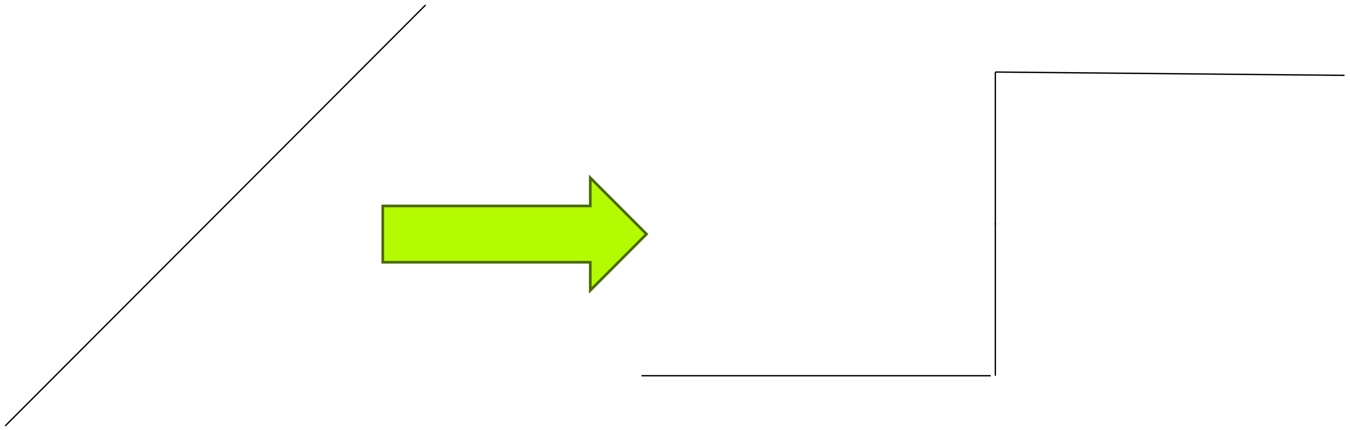
RUN REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin
osarahoittama

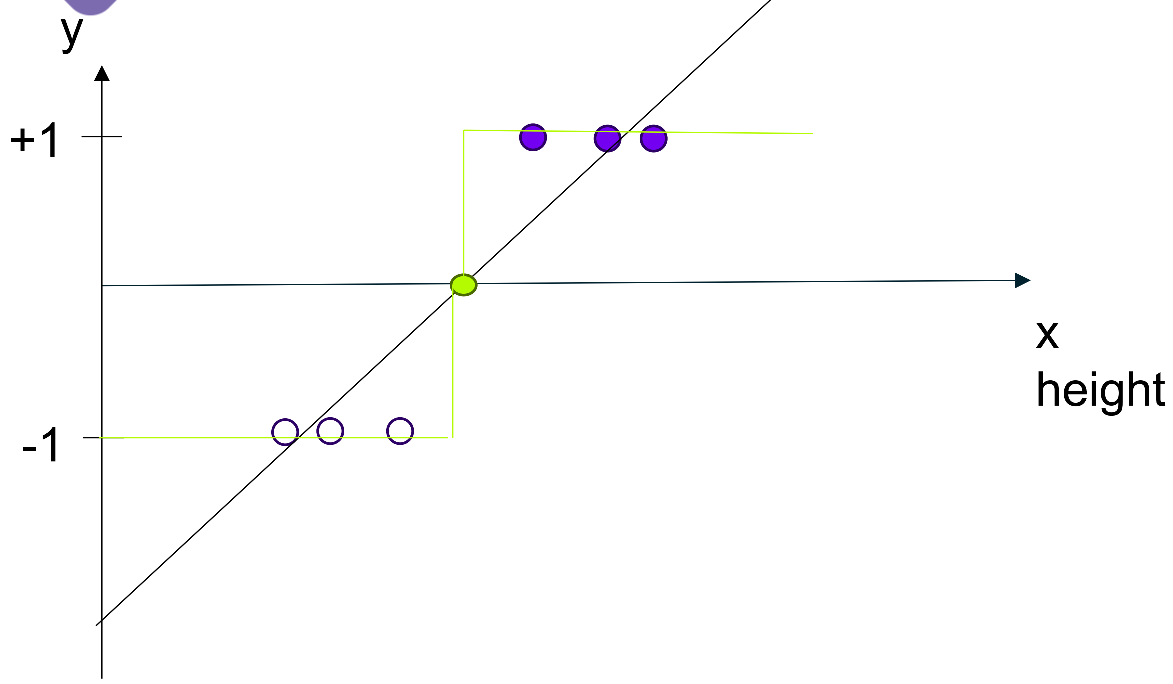
Solving the trouble



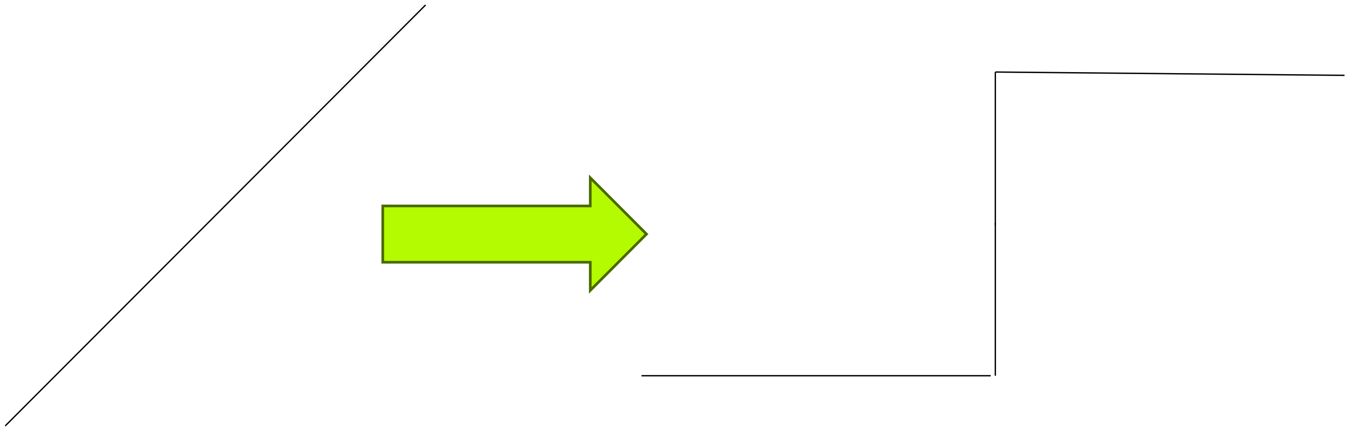
Discretize the problem



Solving the trouble



Discretize the problem



From line to step function or sigmoid function

HAMK
Häme University of Applied Sciences

RUN REGIONAL UNIVERSITY NETWORK EUROPEAN UNIVERSITY

Euroopan unionin osarahoittama

$$S(x) = \frac{1}{1 + e^{-x}}$$

Sigmoid function

x -> -∞

e^{-x} -> ∞

$1 + e^{-x}$ -> ∞

$\frac{1}{1+e^{-x}}$ -> 0

x -> ∞

$\frac{1}{1+e^{-x}}$ -> 1

Two classes: 0 and 1

0: hobbit

1: elf

$$S(x) = \frac{1}{1 + e^{-x}}$$

Classes -1 and +1

-1+2S(x)

S(x) = 0: -1+2*0 = -1

S(x) = 1: -1+2*1 = 1

$$S(x) = \frac{1}{1 + e^{-x}}$$

Where is the line?

- **y=ax+b**
- How to combine sigmoid with y(x)?

HAMK
Häme University
of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin
osarahoittama

$$S(x) = \frac{1}{1 + e^{-x}}$$

Where is the line?

- **y=ax+b**
- How to combine sigmoid with y(x)?
- $\boxed{S(ax+b) = \frac{1}{1+e^{-(ax+b)}}}$
- Meaning of **a** and **b**?

HAMK
Häme University of Applied Sciences

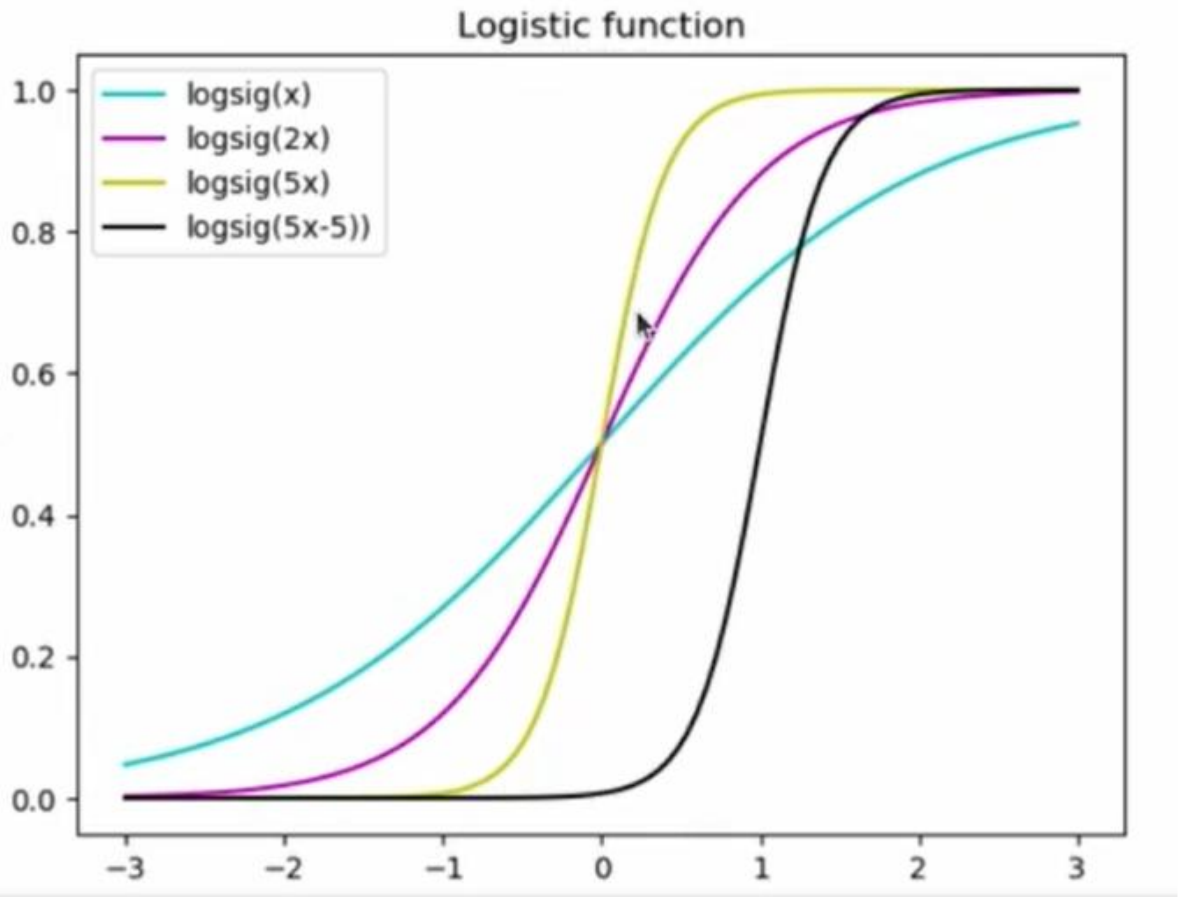
RUN Regional University Network — European University

Euroopan unionin osarahoittama

$$S(x) = \frac{1}{1 + e^{-x}}$$

Where is the line?

- y=ax+b
- How to combine sigmoid with y(x)?
- $\boxed{\mathbf{S(ax+b)} = \frac{1}{1+e^{-(ax+b)}}}$
- Meaning of **a** and **b**?



HAMK
Häme University
of Applied Sciences

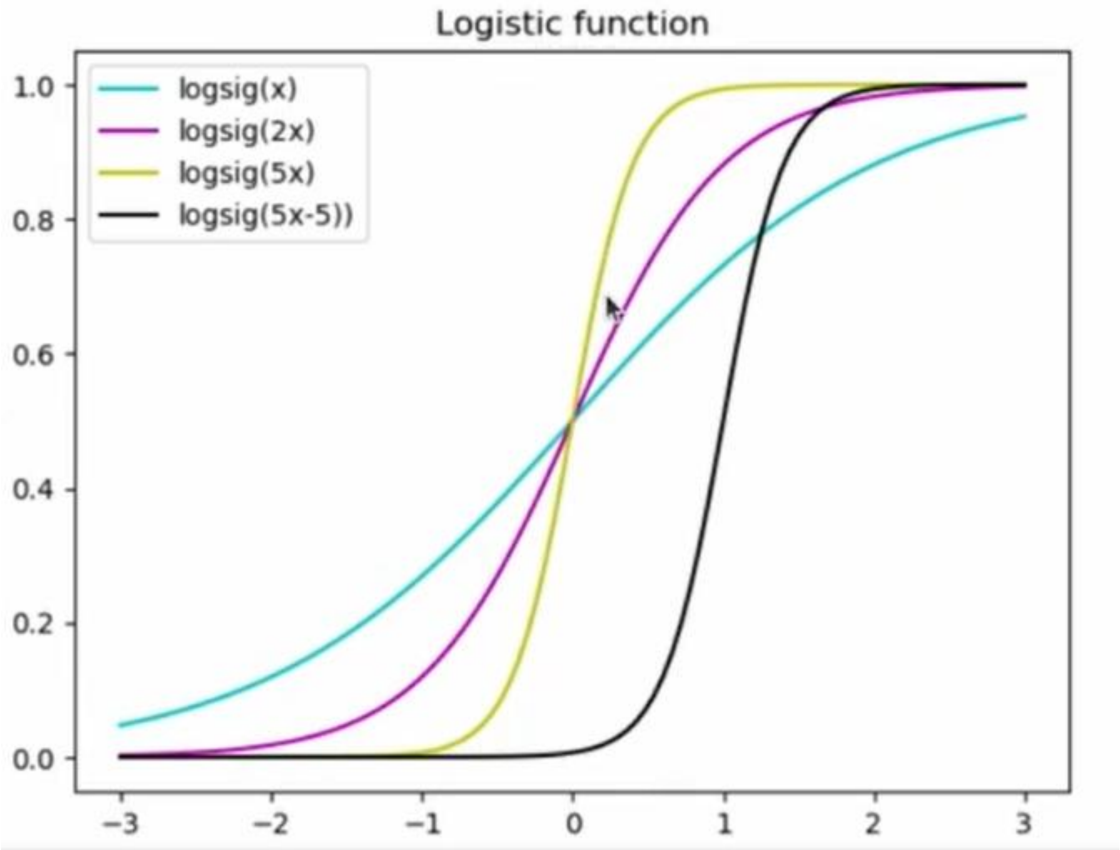
REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin
osarahoittama

$$S(x) = \frac{1}{1+e^{-x}}$$

Where is the line?

- y=ax+b
- How to combine sigmoid with y(x)?
- $\boxed{S(ax+b) = \frac{1}{1+e^{-(ax+b)}}}$
- **Why prefer continuous function rather than step function?**



HAMK
Häme University
of Applied Sciences

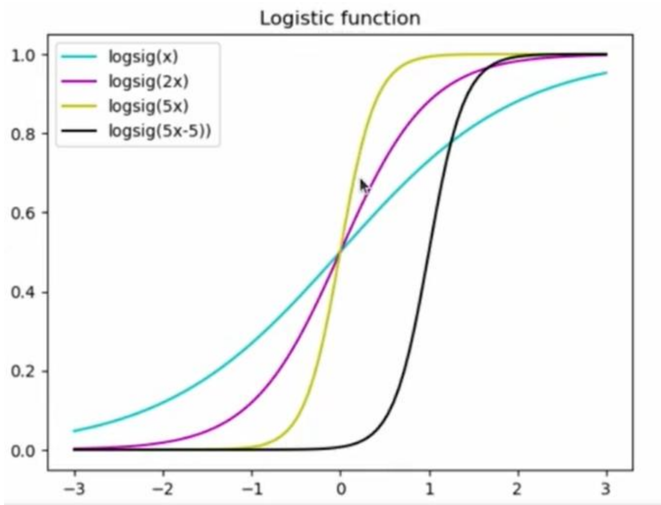
REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin osarahoittama

$$S(x) = \frac{1}{1+e^{-x}}$$

Where is the line?

- y=ax+b
- How to combine sigmoid with y(x)?
- $\boxed{S(ax+b) = \frac{1}{1+e^{-(ax+b)}}}$
- **Why prefer continuous function rather than step function?**
- Step function does not really exist in physical world: it is not continuous, it would require infinite amount of energy to make the sharp step move.



NFL theorem

- Whenever we solve something, we need to pay the price of this lunch.
- Solution: $\frac{\partial\, MSE}{\partial\, a}, \frac{\partial\, MSE}{\partial\, b}$ = 0, 0
- Impossible to solve **a** and **b** (same solution as for deep neural networks).
- MSE = Σ(y_i-y_i_test)^2 =

  = Σ(y_i-S(ax_i_test+b))^2

HAMK
Häme University
of Applied Sciences

REGIONAL UNIVERSITY NETWORK
EUROPEAN UNIVERSITY

Euroopan unionin osarahoittama

Thank you!