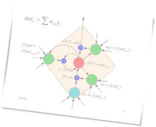
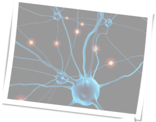
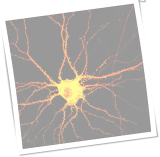
A quick introduction to machine learning

Spyros Samothrakis
Senior Lecturer, IADS
University of Essex
MiSoC



June 22, 2022





Welcome/course contents

- What will this course cover?
  - Day 1: An intro to machine learning (ML)
  - Day 1: ML labs
  - Day 2: An intro to causal inference
  - Day 2: ML and causal inference labs
- Textbooks?
  - Mitchell, T. M. (1997). Machine learning.[1]
  - Bishop, C. M. (2006). Pattern recognition and machine learning. springer.[2]
  - Wasserman, L. (2013). All of statistics: a concise course in statistical inference. Springer Science & Business Media.[3]

[1] `http://www.cs.cmu.edu/~tom/mlbook.html`

[2] `https://www.microsoft.com/en-us/research/publication/pattern-recognition-machine-learning/`

[3] `http://www.stat.cmu.edu/~larry/all-of-statistics/index.html`

Better science through data

Hey, Tony, Stewart Tansley, and Kristin M. Tolle. "Jim Gray on eScience: a transformed scientific method." (2009).[4]

- Thousand years ago: empirical branch
  - You observed stuff and you wrote down about it
- Last few hundred years: theoretical branch
  - Equations of gravity, equations of electromagnetism
- Last few decades: computational branch
  - Modelling at the micro level, observing at the macro level
- Today: data exploration
  - Let machines create models using vast amounts of data

[4] `http://languagelog.ldc.upenn.edu/myl/JimGrayOnE-Science.pdf`

Better business through data

- There was a report by Mckinsey

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.[5]

- Urges everyone to monetise "Big Data"
- Use the data provided within your organisation to gain insights
- Has some numbers as to how much this is worth
- Proposes a number of methods, most of them associated with machine learning and databases

[5] http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Why is it popular now?

- **Algorithms + data + tools**
- Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science, 16(3), 199-231.[6]
- Anderson, P. W. (1972). More is different. Science, 177(4047), 393-396.[7]
- Pedregosa, et.al. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.[8]

[6] http://projecteuclid.org/download/pdf_1/euclid.ss/1009213726%20
[7] https://www.tkm.kit.edu/downloads/TKM1_2011_more_is_different_PWA.pdf
[8] https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

SO THIS COURSE COVERS TOOLS

- ML theory
  - *Supervised learning Regression Classification*
  - Understanding basic modelling
  - Confirming your model is sane
  - Tuning your model
  - **All within a very applied setting**
- Tools
  - Numpy
  - Scikit-learn

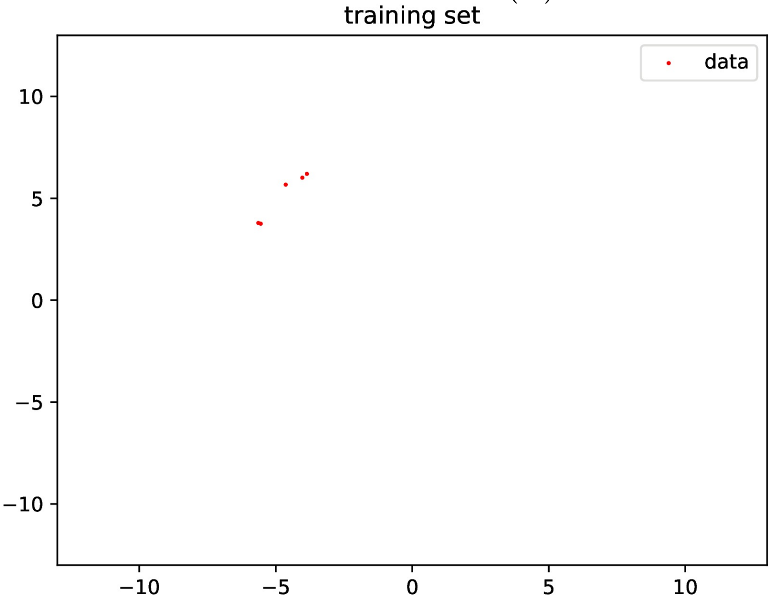
What is supervised learning?

- Imagine someone gives you a group of smokers
  - And asks the question – what is their life expectancy?
- **Completely made up imaginary data**

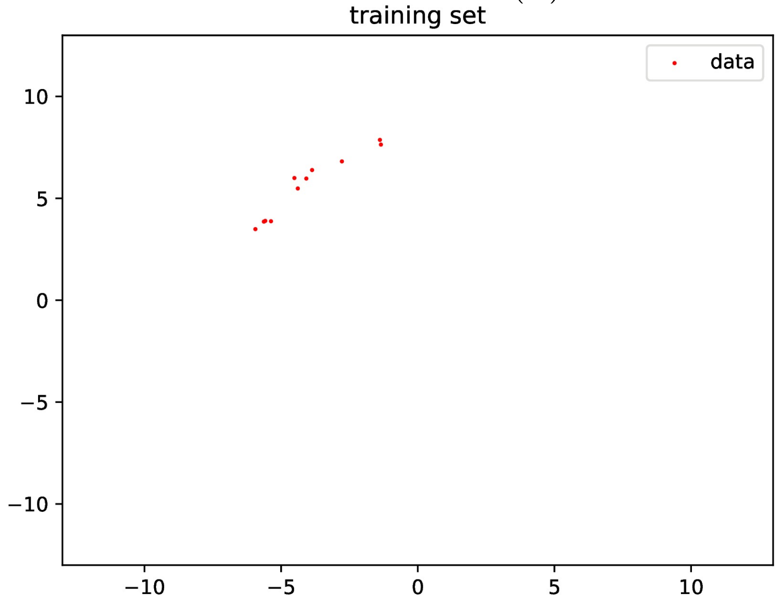
Some abstraction

- We are given inputs $x_0, x_1 ... x_n$ and we are looking to predict y
- Let's plot!

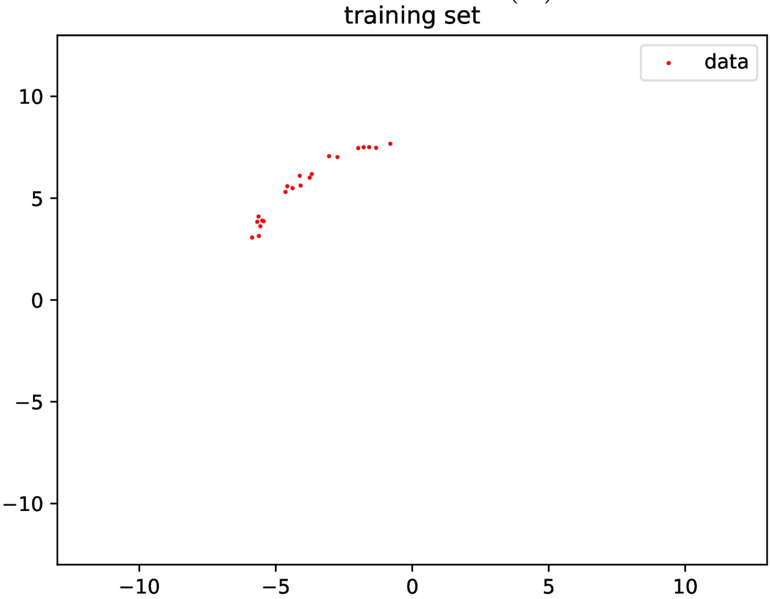
Regression - link the dots (1)



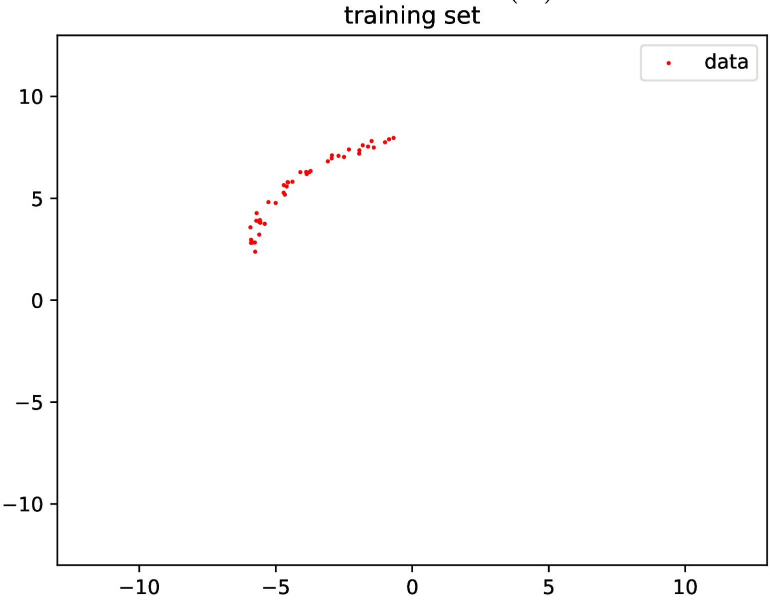
Regression - link the dots (2)



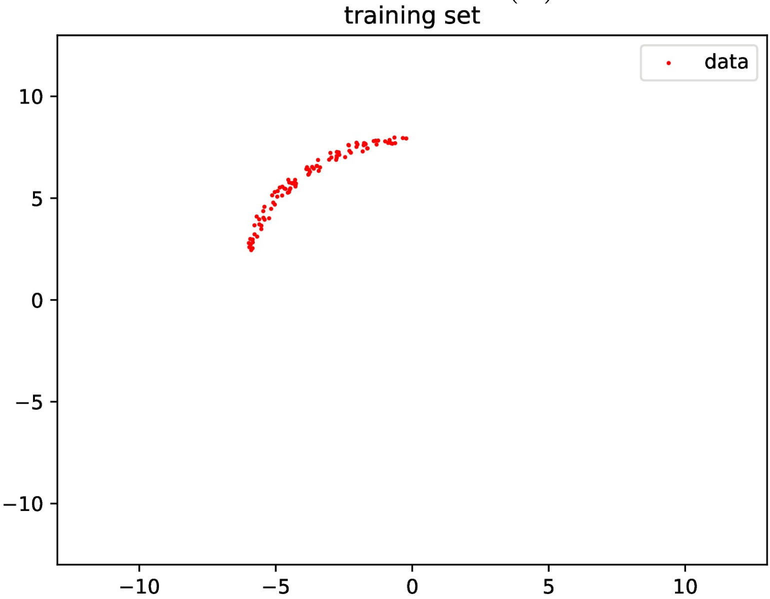
REGRESSION - LINK THE DOTS (3)



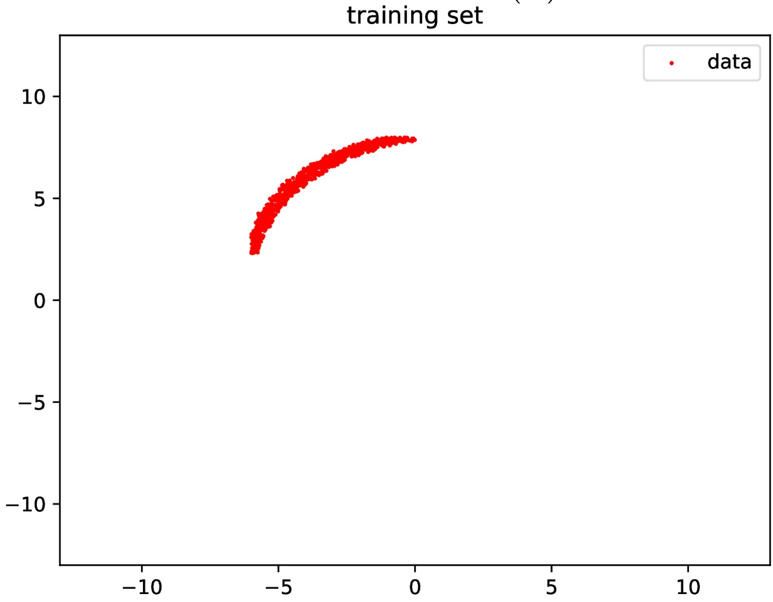
Regression - link the dots (4)



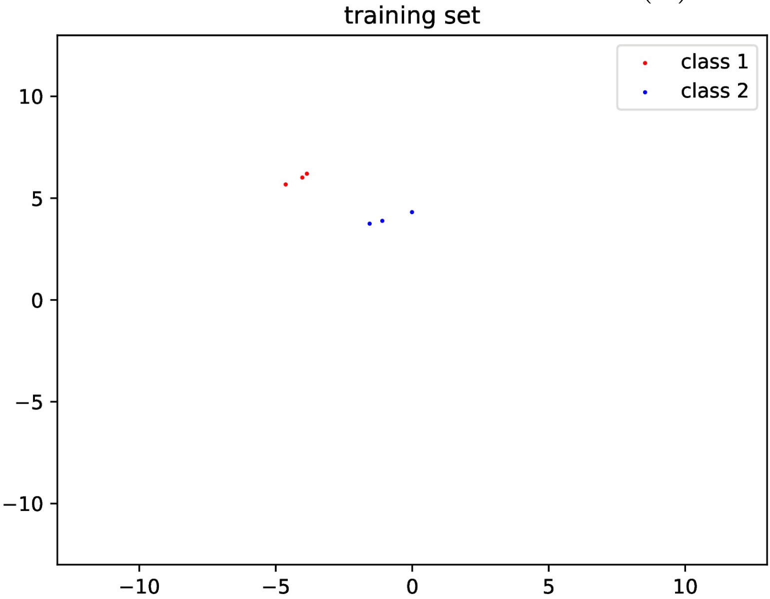
Regression - link the dots (5)



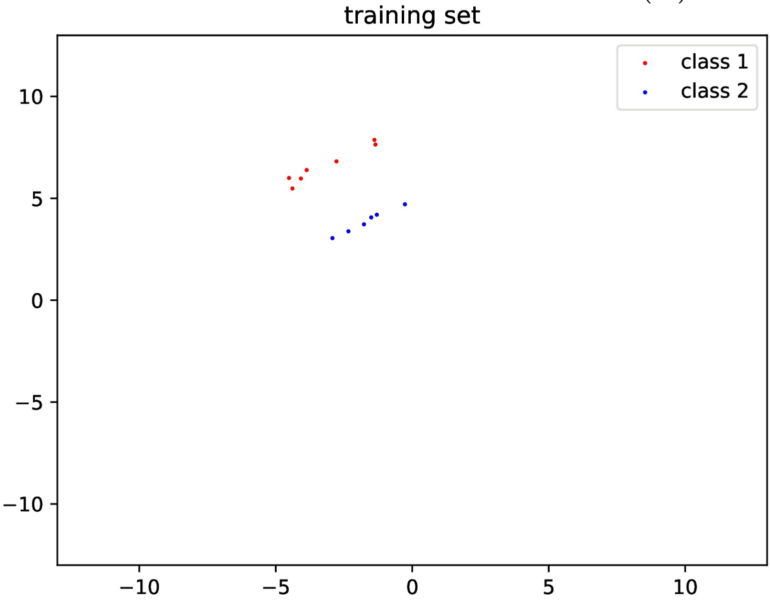
Regression - link the dots (6)



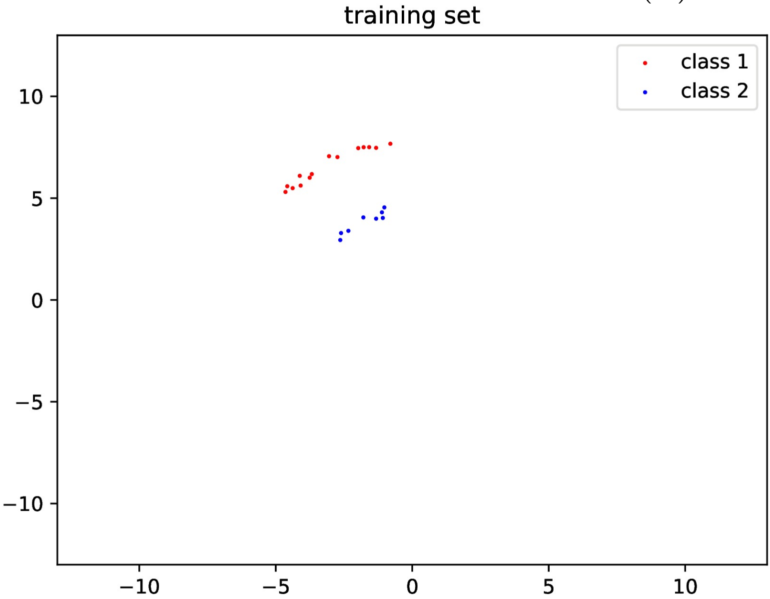
Classification - draw a boundary (1)



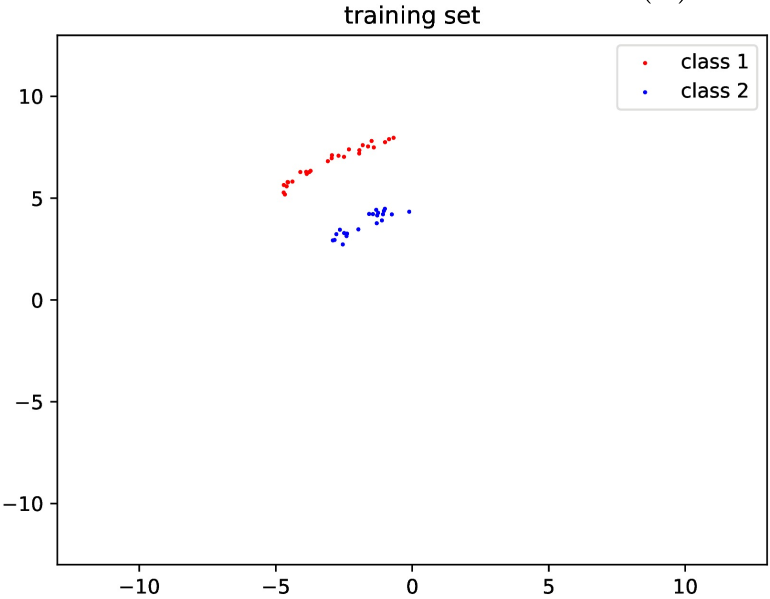
CLASSIFICATION - DRAW A BOUNDARY (2)



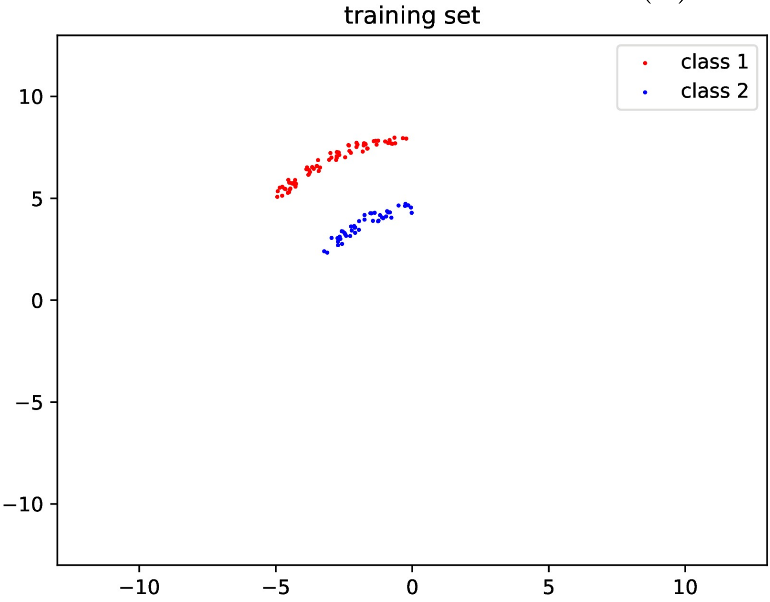
Classification - draw a boundary (3)



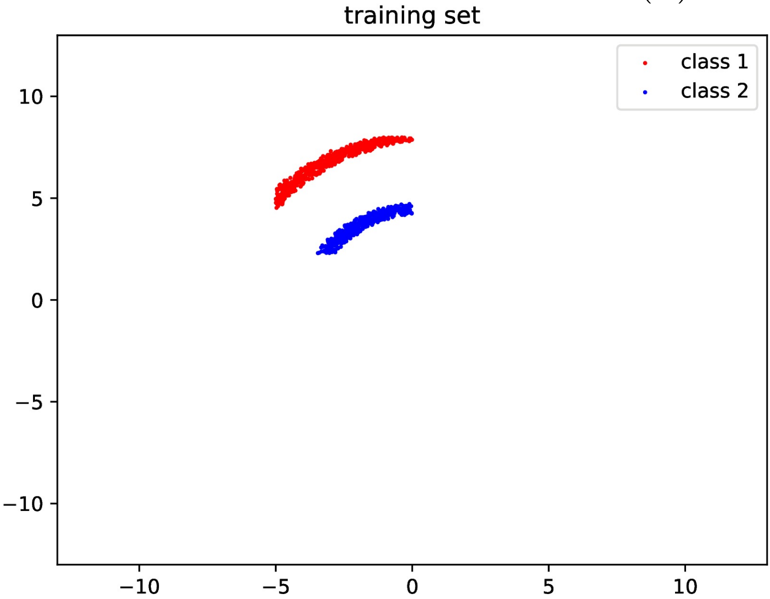
Classification - draw a boundary (4)



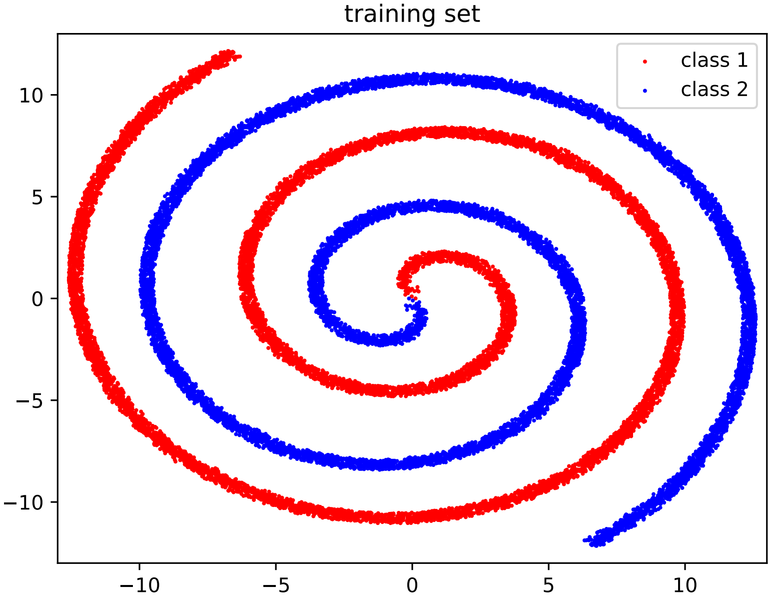
Classification - draw a boundary (5)



Classification - draw a boundary (6)



Full data



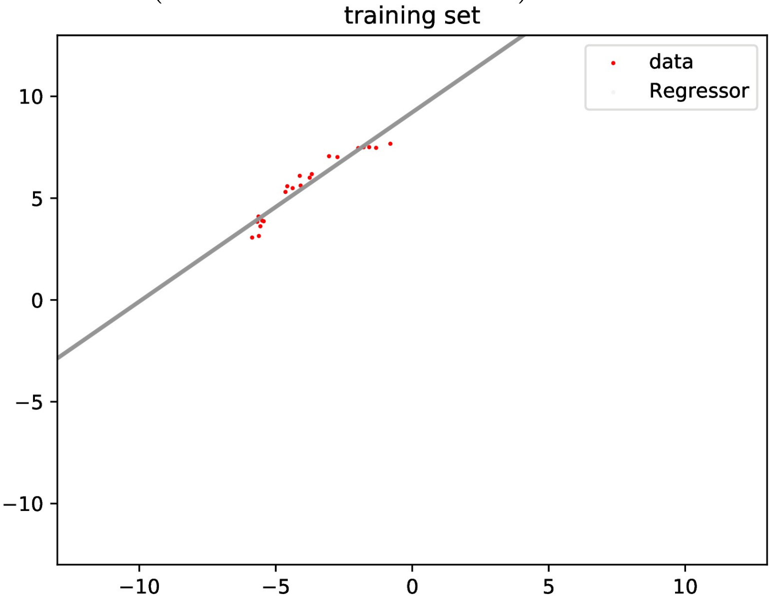
Intuition

- That's it - we are given data, and we need to come up with an algorithm to join it up – but in high dimensions
  - Can can be binary, categorical, real-valued
- How well well a function joins the data is called the "loss"
- Very low loss is not good, it might not generalise that well to unseen data points – you can learn to memorise data instances

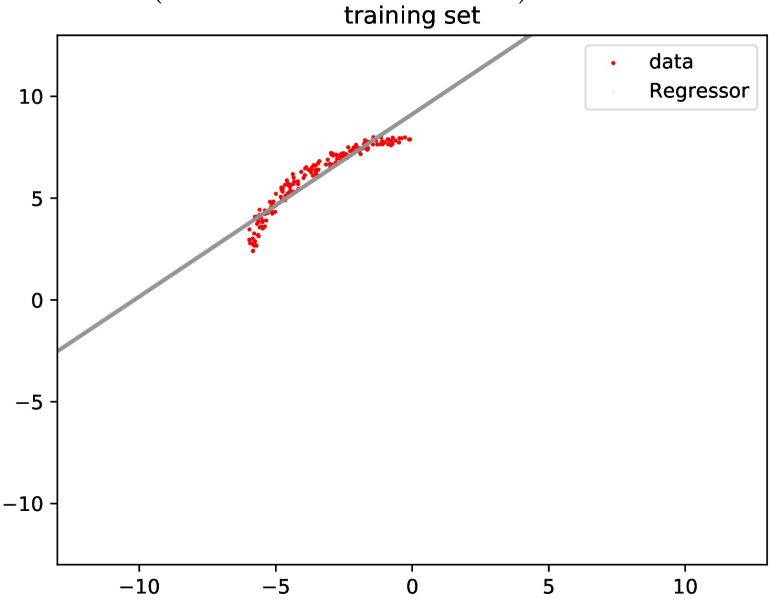
Linear regression

- Linear and logistic regression
  - Logistic regression does classification
- You just assume everything is a line

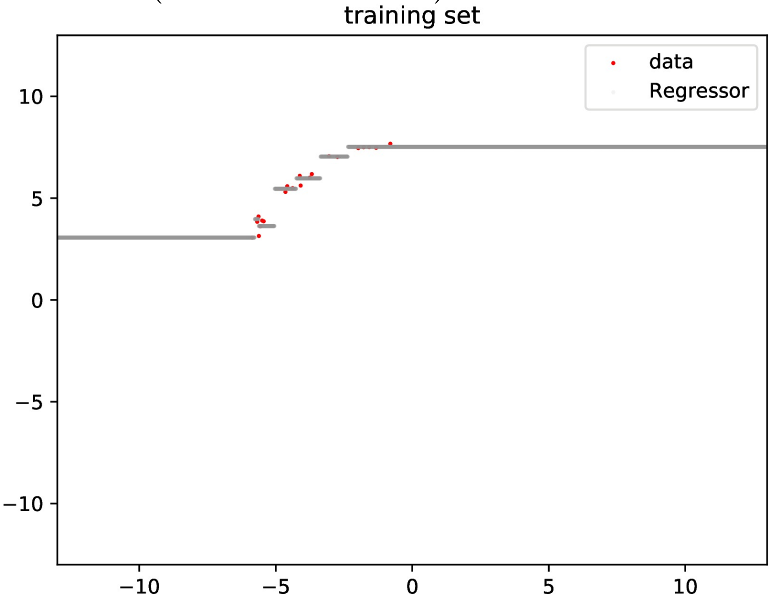
Example (Linear regression)



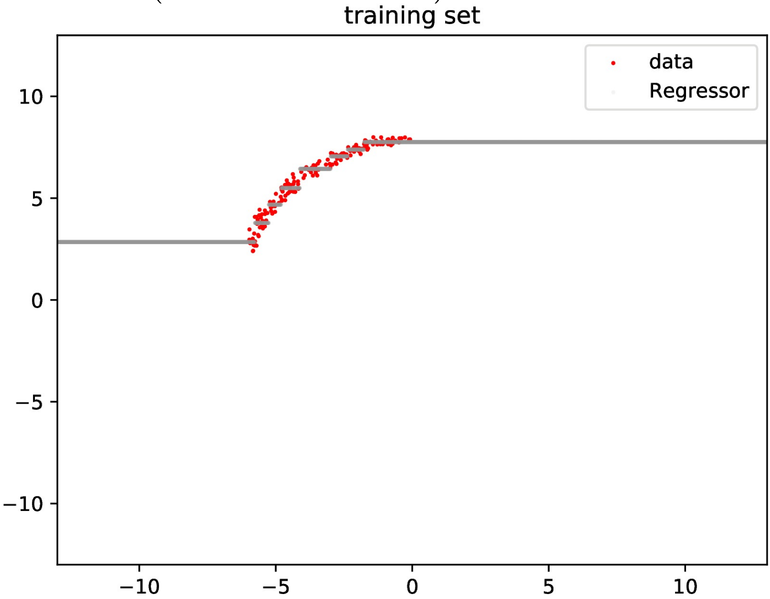
Example (Linear regression)



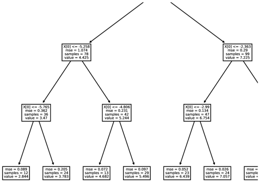
Example (Decision tree)



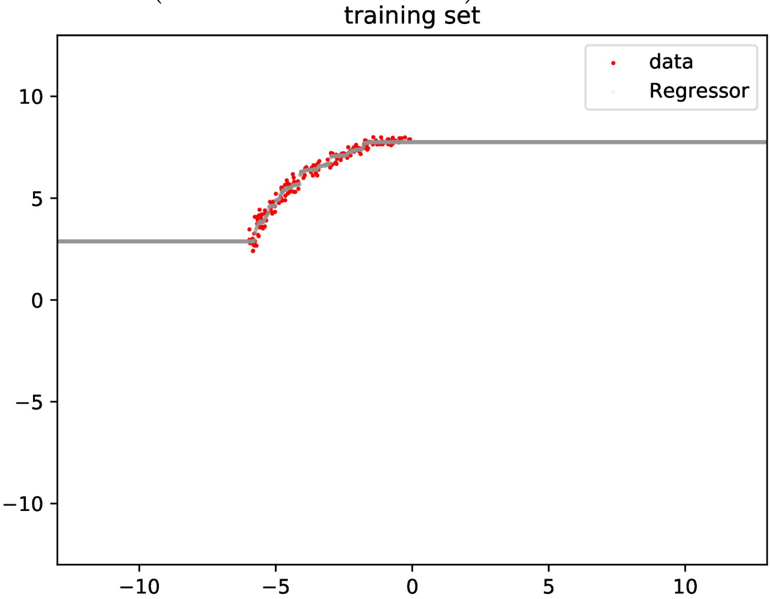
Example (Decision tree)



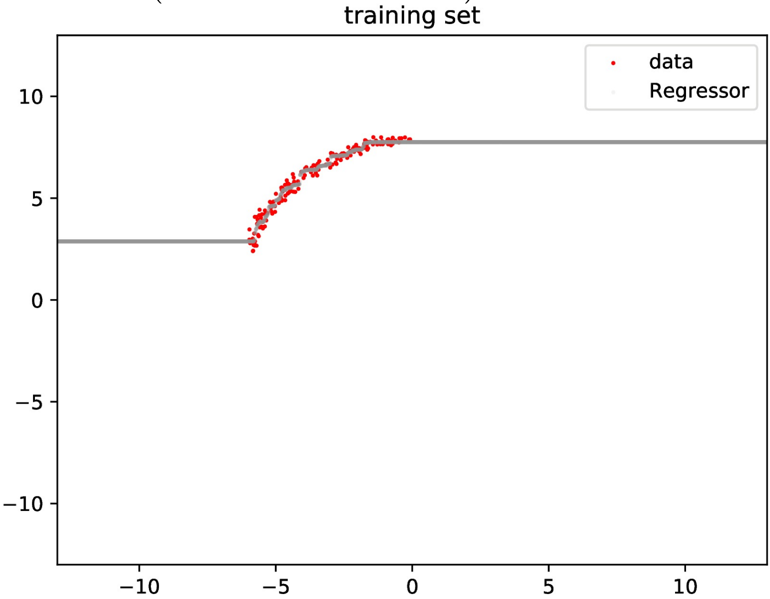
Example (Decision tree — internal)



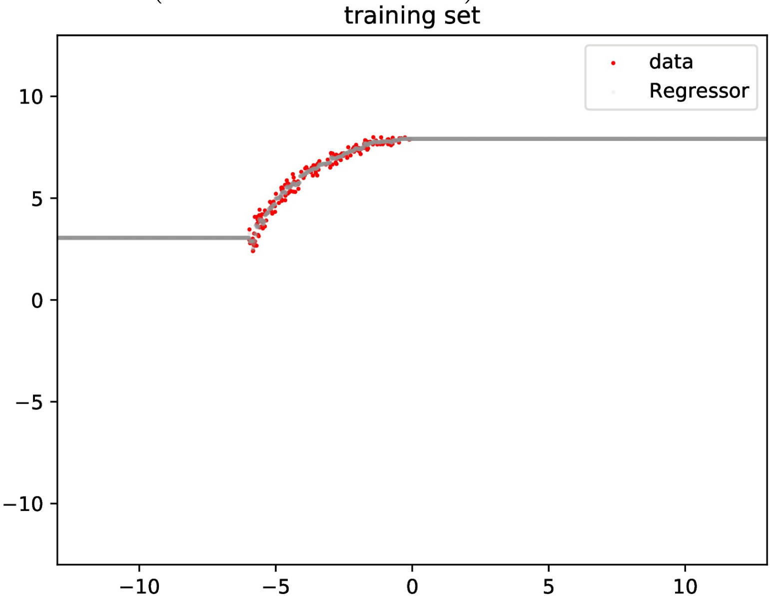
EXAMPLE (RANDOM FOREST)



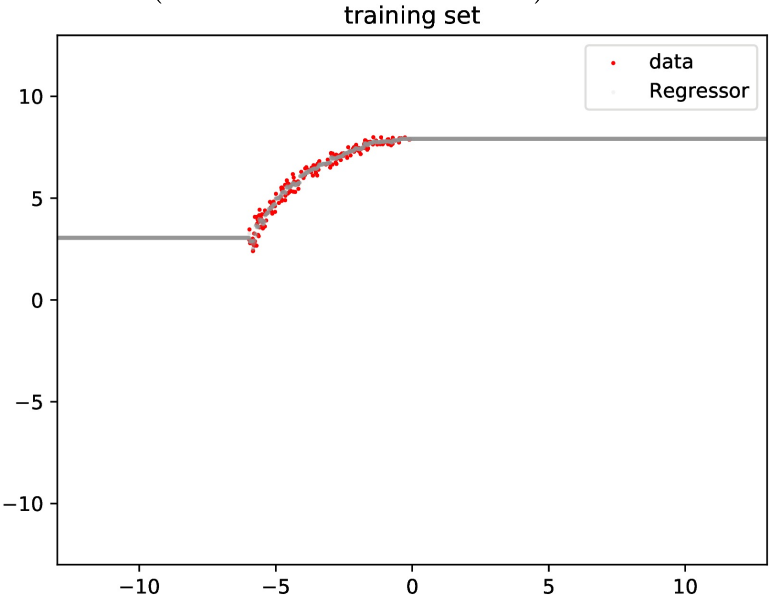
Example (Random forest)



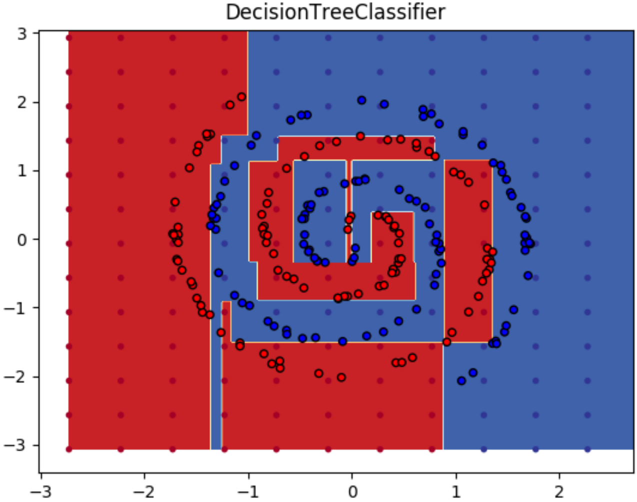
Example (Random forest)



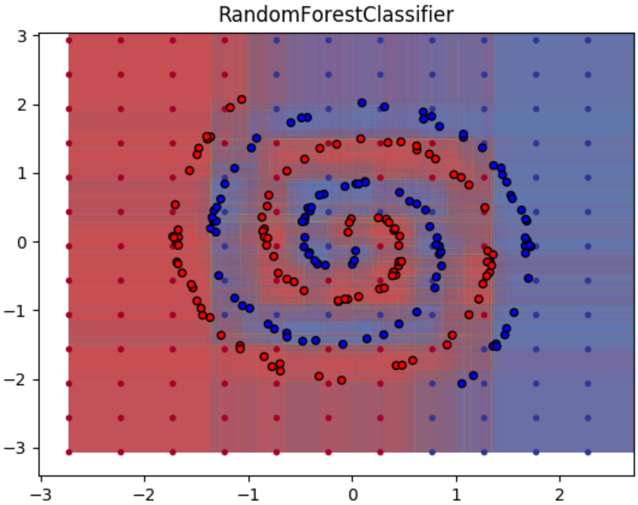
Example (Gradient boosting)



Classification (Decision trees)



Classification (Random forests)



Until now

But how do we know this will generalise well?

- Train/Validation/Test split
- Cross validation

Hyperparameters

- How many trees?
- Tree depth?
- l2?