Machine Learning Concepts

BSDS 100, Spring 2021
Michael Ruddy

What is Machine Learning?

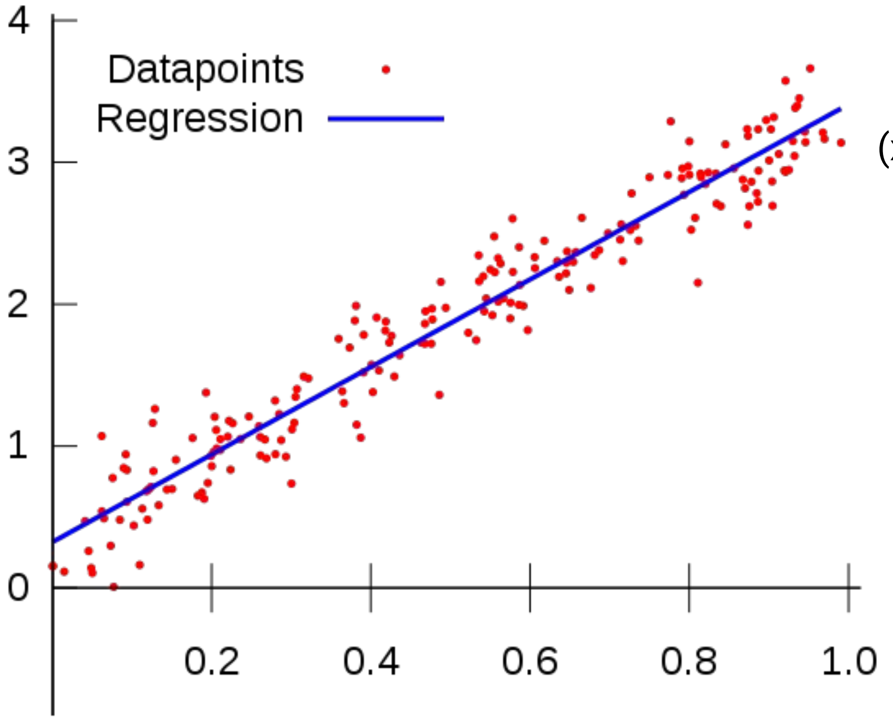
- Broadly, machine learning consists of creating and using models trained on data (or *learned* from data).
    - Study of computer algorithms that automatically take in data and output high-performing models
    - Often seen as part of artificial intelligence

Supervised vs. Unsupervised Learning

- **Supervised Learning:** Learn a model based on a dataset on input/output pairs, i.e. *labelled* data.
  - Examples:
    - Linear Regression
    - Deep Learning
- **Unsupervised Learning:** Learn patterns in *unlabelled* data.
  - Examples
    - Clustering Algorithms
    - Anomaly Detection

Supervised Learning

- Learn how to predict Y from X
- Linear Regression:
  y=mx+b
- "Find values for the parameters m and b using (x,y) points," i.e. from labelled data



(x,y)

Photo from [here]

Supervised Learning

- Deep Learning is often supervised learning

Data

- Find parameters for a *neural network* to predict either 1 or 0 ("cat" or "not a cat") from an input image.

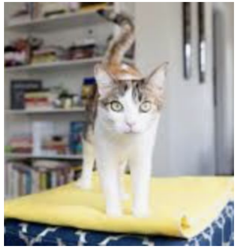
Cat

Cat

Not a Cat

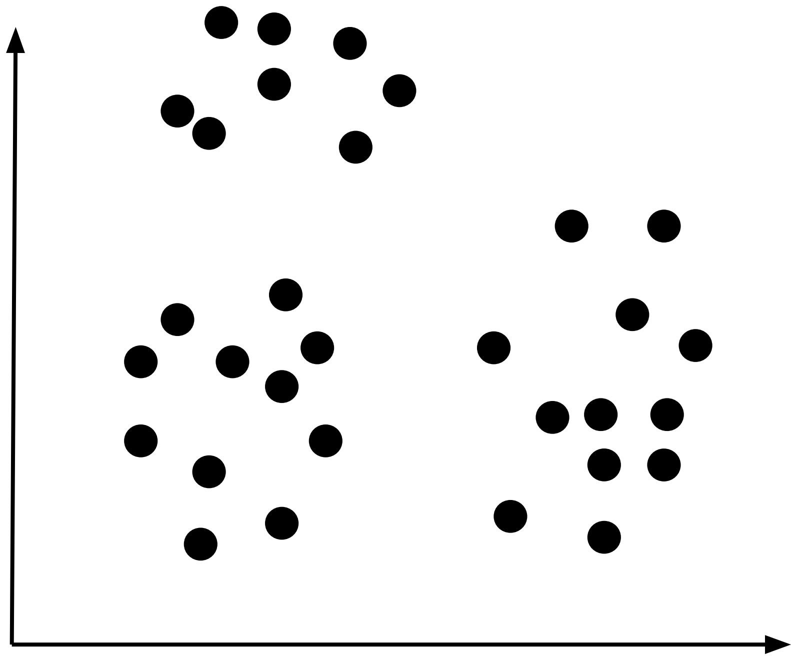
Cat

Not a Cat

$f(\vec{a},$  $) = .946$

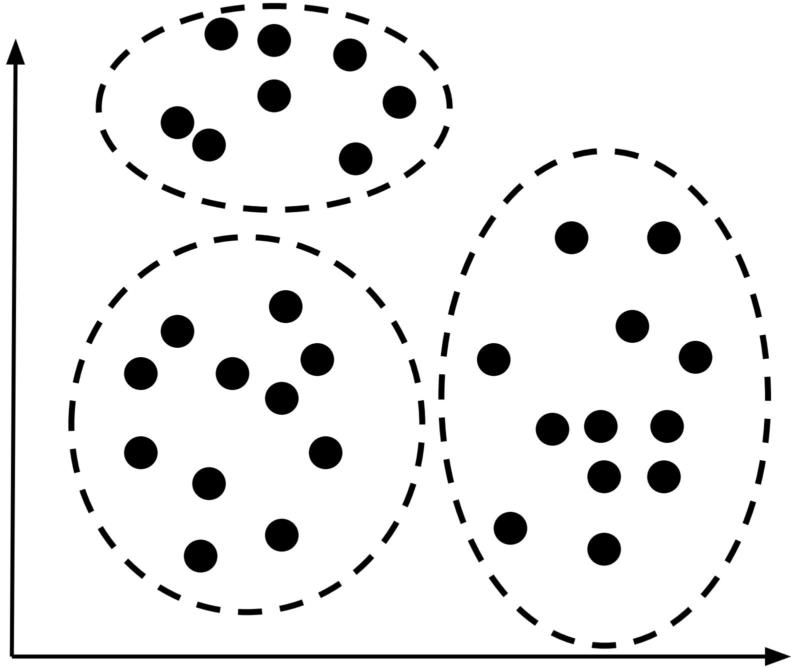
Unsupervised Learning

- Understand patterns in *unlabelled* data
- For example, Clustering algorithms take unlabelled data and arrange them into groups or clusters
- Up to human to assign meaning to clusters



Unsupervised Learning

- Understand patterns in *unlabelled* data
- For example, Clustering algorithms take unlabelled data and arrange them into groups or clusters
- Up to human to assign meaning to clusters



Self-Supervised Learning

April 26

From Wikipedia, the free encyclopedia

April 26 is the 116th day of the year (117th in leap years) in the Gregorian calendar; 249 days remain until the end of the year.

- A “fake” supervised learning task when your desired task does not have a lot of data
- Natural Language Processing: Algorithms for processing and analyzing human language

Self-Supervised Learning

April 26

From Wikipedia, the free encyclopedia

?

April 26 is the 116th day of the year (117th in leap years) in the Gregorian calendar; 249 days remain until the end of the [redacted].

- A “fake” supervised learning task when your desired task does not have a lot of data
- Natural Language Processing: Algorithms for processing and analyzing human language
- Predict next word

Self-Supervised Learning

April 26

From Wikipedia, the free encyclopedia

?

April 26 is the 116th day of the year (117th in leap years) in the Gregorian calendar; 249 days remain until the end of the [redacted].

- A “fake” supervised learning task when your desired task does not have a lot of data
- Natural Language Processing: Algorithms for processing and analyzing human language
- Predict next word -> Create model for translation

Self-Supervised Learning

April 26

From Wikipedia, the free encyclopedia

?

April 26 is the 116th day of the year (117th in leap years) in the Gregorian calendar; 249 days remain until the end of the [redacted].

- A “fake” supervised learning task when your desired task does not have a lot of data
- Natural Language Processing: Algorithms for processing and analyzing human language
- Predict next word -> Create model for translation
- (Input/Output machine w/ parameters chosen using data!)

Loss Function for Supervised Learning

- For supervised learning, the **Loss Function** (or **Cost Function**) is the function which determines how well a model fits the data.
- An **Optimization Algorithm** is used to find the *parameters* which give the lowest score on the training data
    - Gradient Descent, Newton’s Method (these are basic calculus!)

Model Parameters

Input Label

Loss Function

$$\mathcal{L}(f(\vec{a}, X), Y)$$

Linear Regression

- Data: $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$
- Model: $y = a_1 + a_2 x$
- Loss function: Root Mean Squared Error

$$\sqrt{\frac{(y_1 - (a_1 + a_2 x_1))^2 + \cdots + (y_N - (a_1 + a_2 x_N))^2}{N}}$$

Linear Regression

- Data: $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$
- Model: $y = a_1 + a_2 x \longleftarrow f(\vec{a}, x)$
- Loss function: Root Mean Squared Error

$$\mathcal{L}(f(\vec{a}, x), y) = \sqrt{\frac{(y_1 - (a_1 + a_2 x_1))^2 + \cdots + (y_N - (a_1 + a_2 x_N))^2}{N}}$$

- Find the value of $\vec{a}$ that minimizes the loss.

Supervised Learning

- Ingredients:
  - Model Family
  - Loss Function
  - Method to choose best Model
  - Lots and lots of data

Supervised Learning

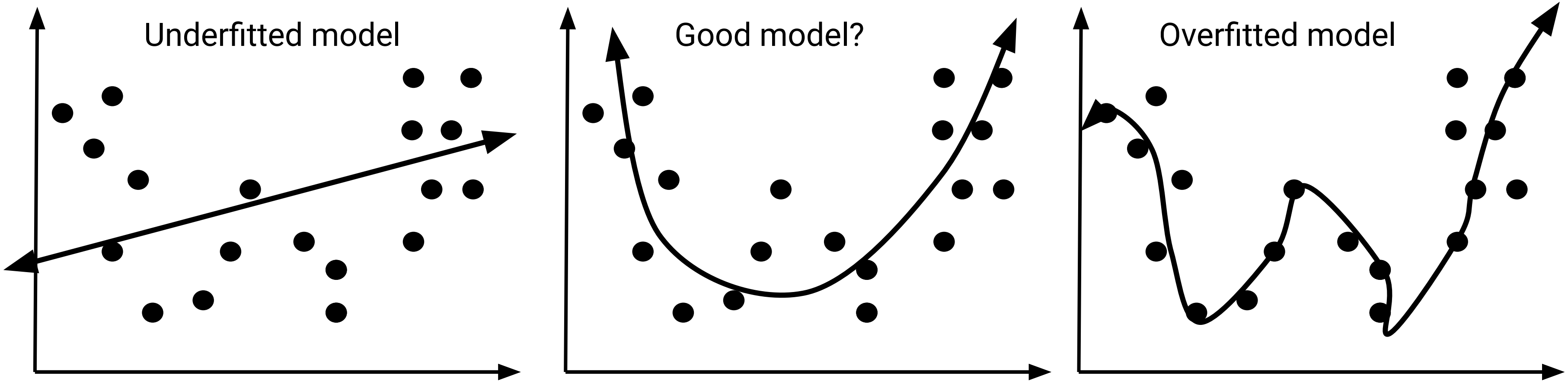
- Ingredients:
    - Model Family
    - Loss Function
    - Method to choose best Model
    - Lots and lots of data *****

Underfitting vs. Overfitting

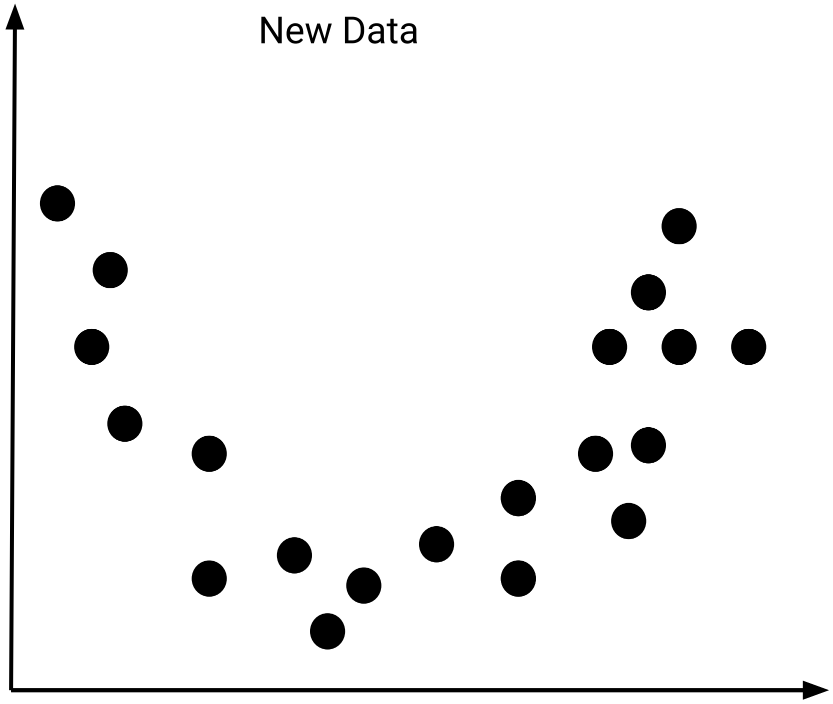
- Recall that we use a Train/Valid/Test split to develop models that **generalize well** to unseen data.
- The model with the best error score isn't always the "best" model!

Underfitting vs. Overfitting

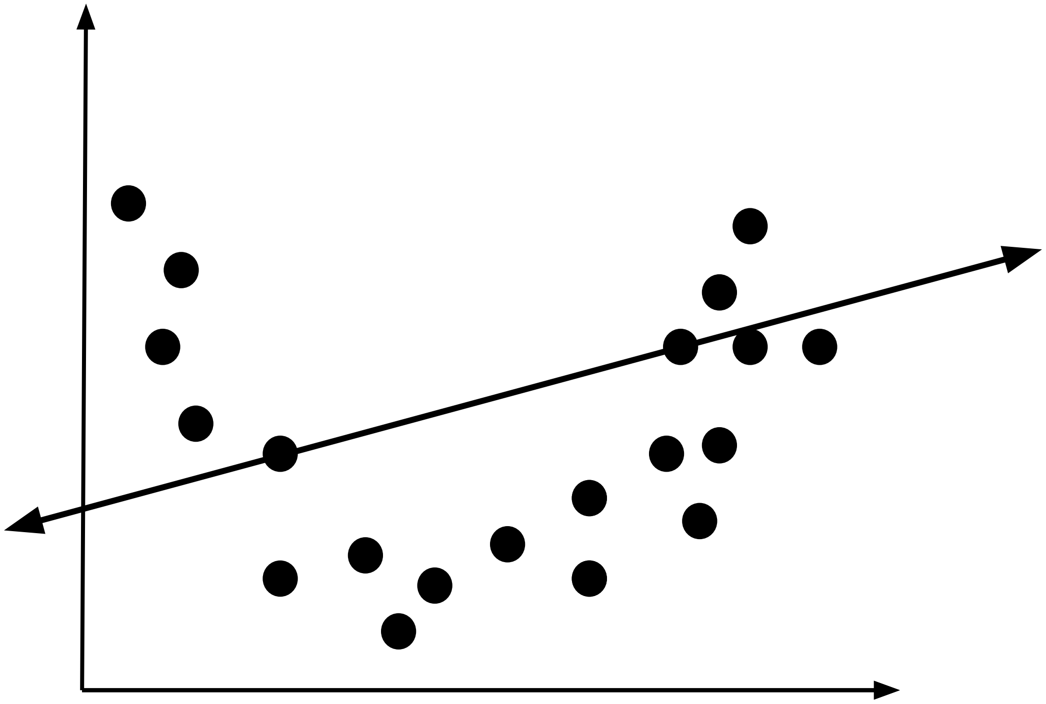
- Recall that we use a Train/Valid/Test split to develop models that **generalize well** to unseen data.
- The model with the best error score isn't always the "best" model!



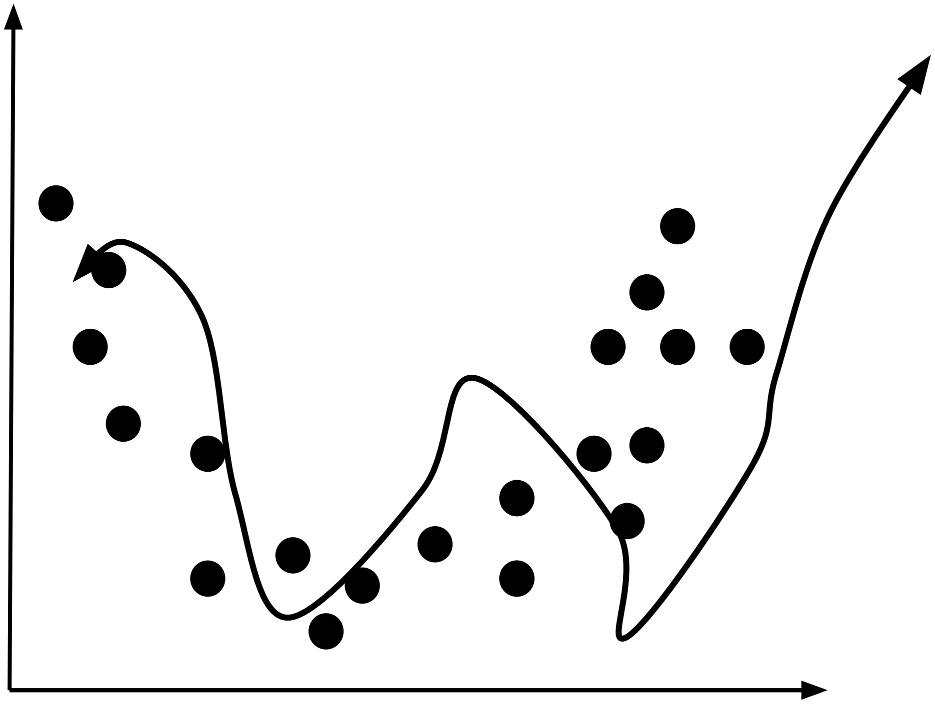
Underfitting vs. Overfitting



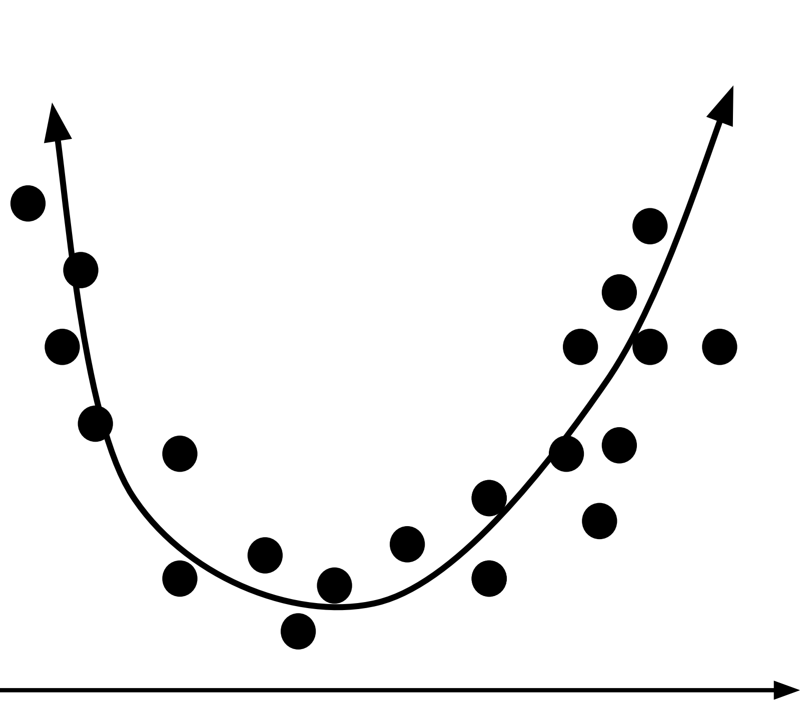
Underfitting vs. Overfitting



Underfitting vs. Overfitting



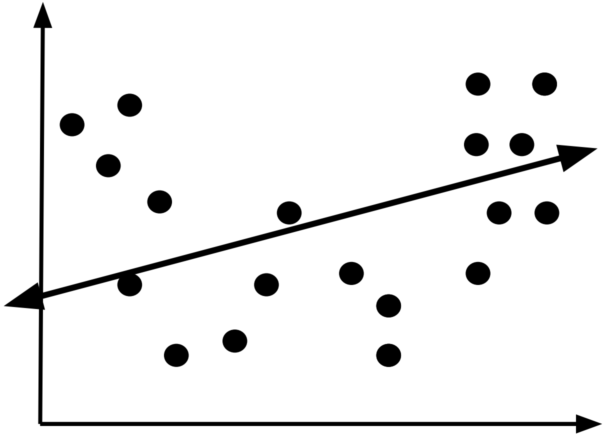
Underfitting vs. Overfitting



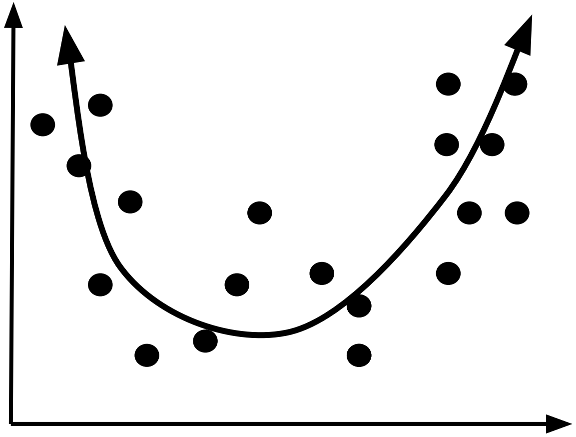
Bias-Variance Tradeoff

- *Bias* error results from a model not predicting well
- *Variance* error results from predictions being sensitive to perturbations

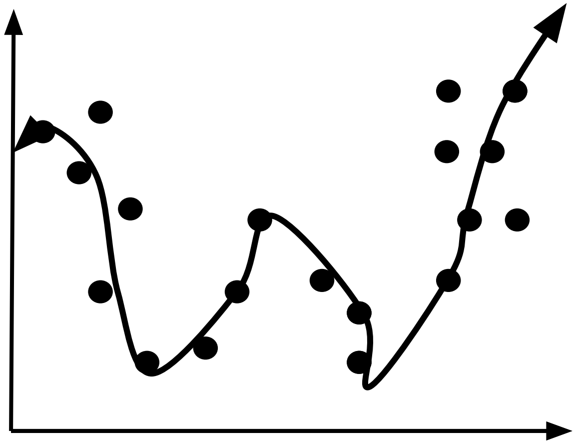
High Bias
Low Variance



Acceptable Bias
Acceptable Variance



Low Bias
High Variance



Bias-Variance Tradeoff

- How to reduce Variance (and overfitting)
  - Use simple model families (i.e. linear models)
  - Add penalty terms to the Loss Function (measure of variance)
  - More data
- How to reduce Bias (and underfitting)
  - Use more complex model families
  - More data

Metric vs. Loss Function

- A **metric** is the “human-desired” measure of how well your model performs, while the loss function is a “mathematical” score used for optimization.
- Example: Binary Classification for Cat vs. Non-Cat
    - Goal: Output probability input is a Cat
    - Loss Function: Likelihood function
    - Metric: Accuracy

Summary

- Supervised learning is sometimes as simple as finding the best parameters for your model family according to your training data
- Have to use techniques to ensure good generalization
- Falls under Artificial Intelligence: “Automatically” construct a good model just by feeding it lots of data (not always human interpretable!)
  - Deep Learning is an extreme example of this
  - Complex model structure, needs lots of data