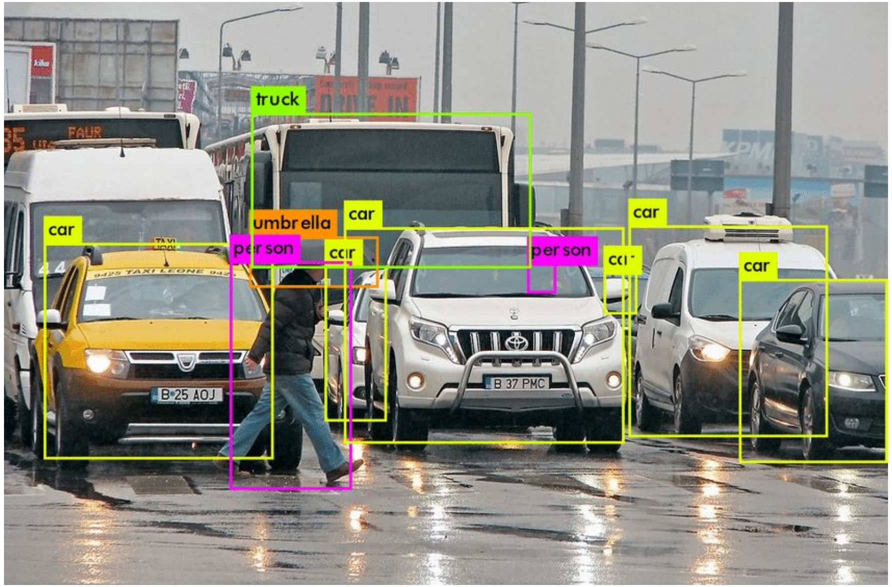
Dataset Preparation and Result Analysis

Lab 02

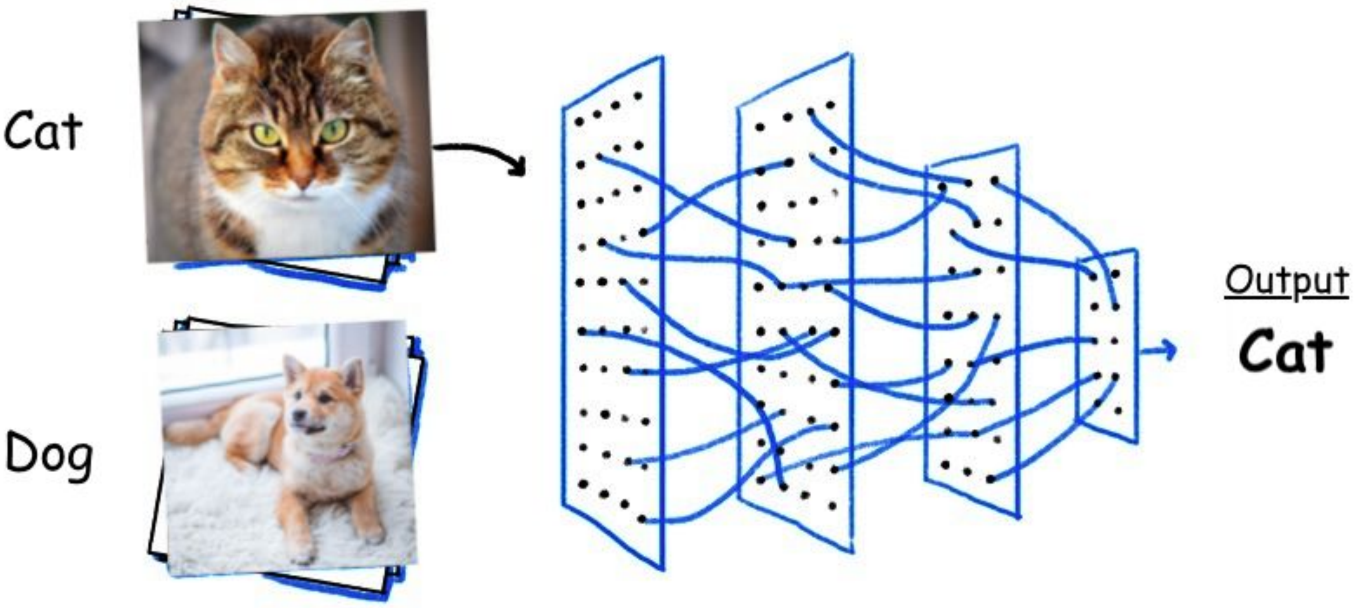
How to Make Dataset

- ➔ Pre-Existing Dataset (eg. Kaggle)
- ➔ Make your own but remember
  1) Angle
  2) Light
  3) Distance
  4) Noise
- ➔ Data augmentation
- ➔ Data Labelling
- ➔ Ground Truth

Ground Truth



Data Labelling



Confusion Matrix

What is Confusion Matrix and why you need it?

Well, it is a performance measurement for machine learning classification problem where output can be two or more classes. It is a table with 4 different combinations of predicted and actual values.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Cont.

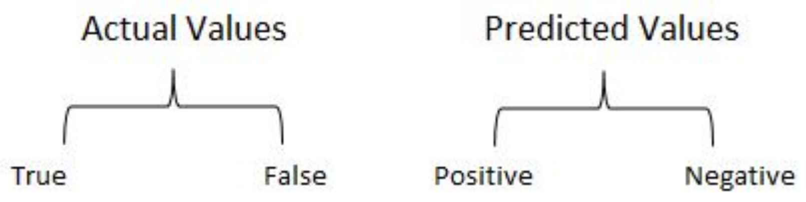
True Positive: Interpretation: You predicted positive and it's true.You predicted that a woman is pregnant and she actually is.

True Negative: Interpretation: You predicted negative and it's true.You predicted that a man is not pregnant and he actually is not.

False Positive: (Type 1 Error) Interpretation: You predicted positive and it's false. You predicted that a man is pregnant but he actually is not.

False Negative: (Type 2 Error) Interpretation: You predicted negative and it's false. You predicted that a woman is not pregnant but she actually is.

Just Remember, We describe predicted values as Positive and Negative and actual values as True and False.



Recall

Out of all the positive classes, how much we predicted correctly. It should be high as possible.

$$\textit{Recall} = \frac{TP}{TP + FN}$$

Precision

Out of all the positive classes we have predicted correctly, how many are actually positive.

$$\textbf{\textit{Precision}} = \frac{TP}{TP + FP}$$

F-measure

It is difficult to compare two models with low precision and high recall or vice versa. So to make them comparable, we use F-Score. F-score helps to measure Recall and Precision at the same time. It uses Harmonic Mean in place of Arithmetic Mean by punishing the extreme values more.

$$\textit{F - measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

Accuracy

Accuracy is one metric for evaluating classification models. Informally, **accuracy** is the fraction of predictions our model got right. Formally, accuracy has the following definition:

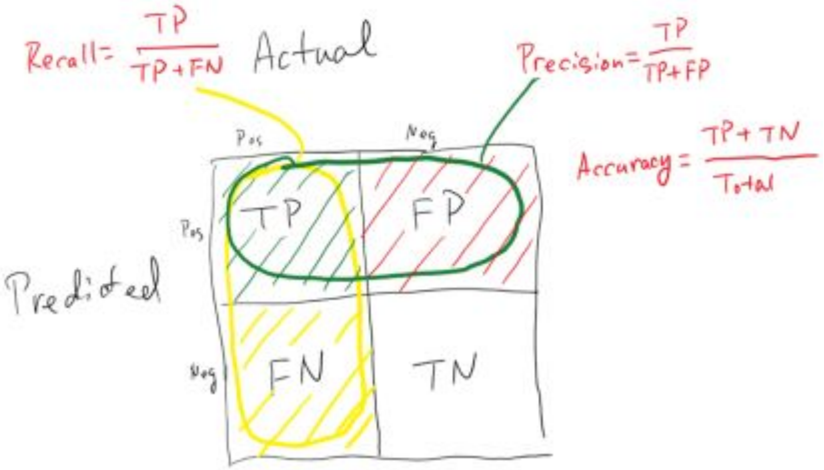
$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

For binary classification, accuracy can also be calculated in terms of positives and negatives as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Class Task

y	y pred	output for threshold 0.6
0	0.5	0
1	0.9	1
0	0.7	1
1	0.7	1
1	0.3	0
0	0.4	0
1	0.5	0



Answer

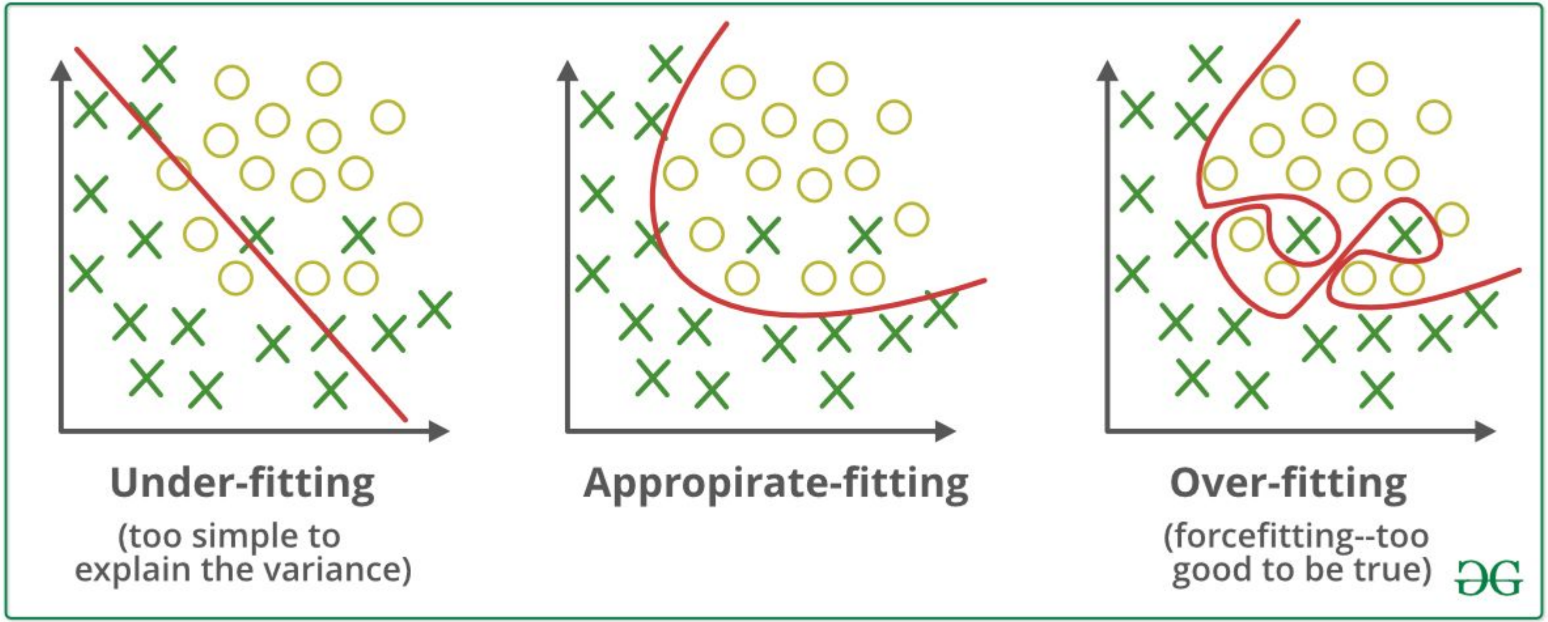
Actual

Predict

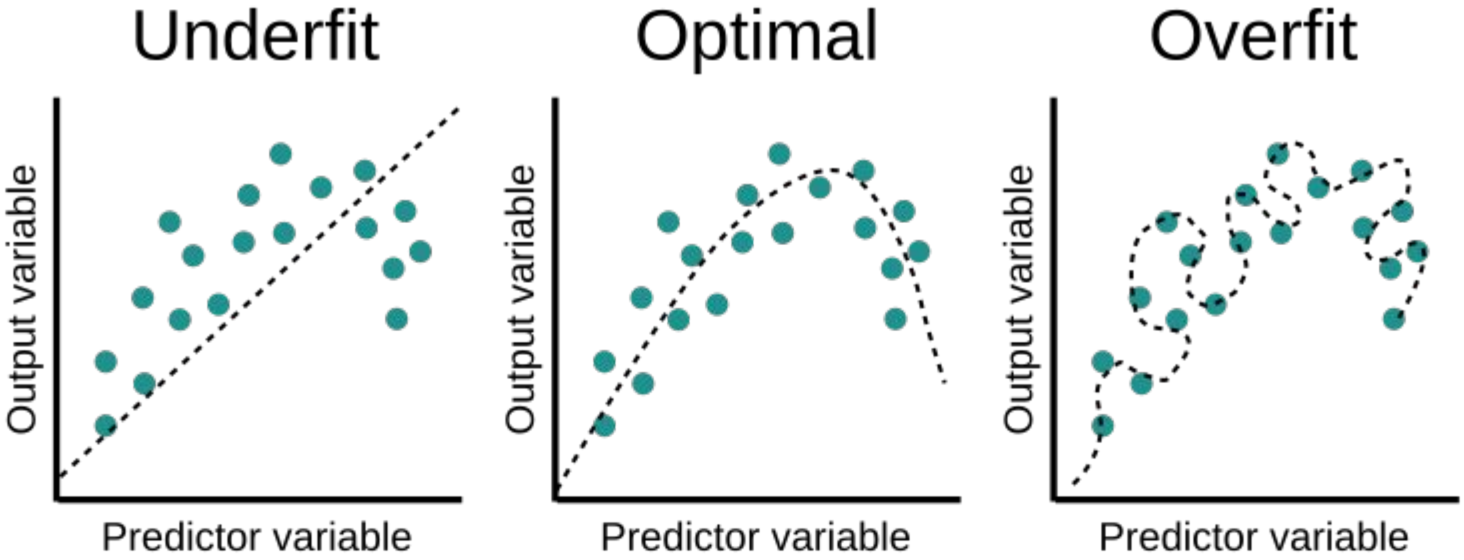
TP =2	FP=1
FN=2	TN=2

<table>
<tr><th>y</th><th>y pred</th><th>output for threshold 0.6</th><th>Recall</th><th>Precision</th><th>Accuracy</th></tr>
<tr><td>0</td><td>0.5</td><td>0</td><td rowspan="7">1/2</td><td rowspan="7">2/3</td><td rowspan="7">4/7</td></tr>
<tr><td>1</td><td>0.9</td><td>1</td></tr>
<tr><td>0</td><td>0.7</td><td>1</td></tr>
<tr><td>1</td><td>0.7</td><td>1</td></tr>
<tr><td>1</td><td>0.3</td><td>0</td></tr>
<tr><td>0</td><td>0.4</td><td>0</td></tr>
<tr><td>1</td><td>0.5</td><td>0</td></tr>
</table>

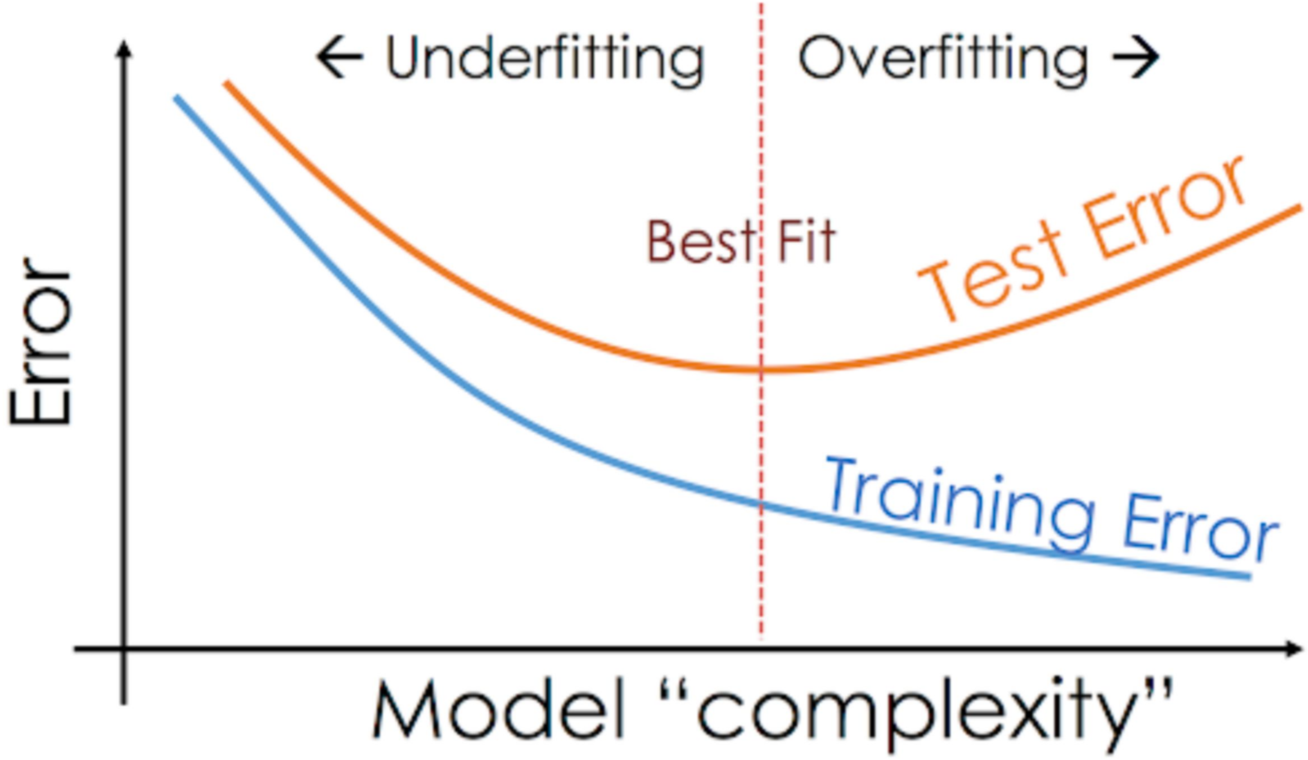
cont.


Under-fitting
(too simple to
explain the variance)
Appropirate-fitting
Over-fitting
(forcefitting--too
good to be true)

Cont.


Underfit
Optimal
Overfit
Output variable
Output variable
Output variable
Predictor variable
Predictor variable
Predictor variable

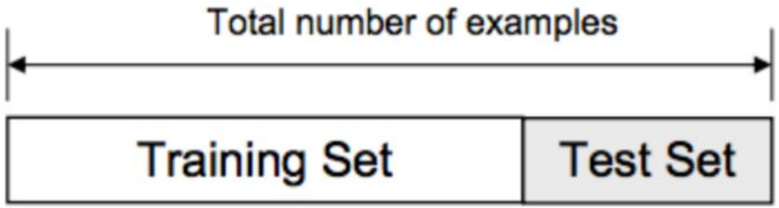
Cont.



← Underfitting
Overfitting →
Best Fit
Test Error
Training Error
Error
Model "complexity"

Data Validation

- **Train/Validation/Test Split**

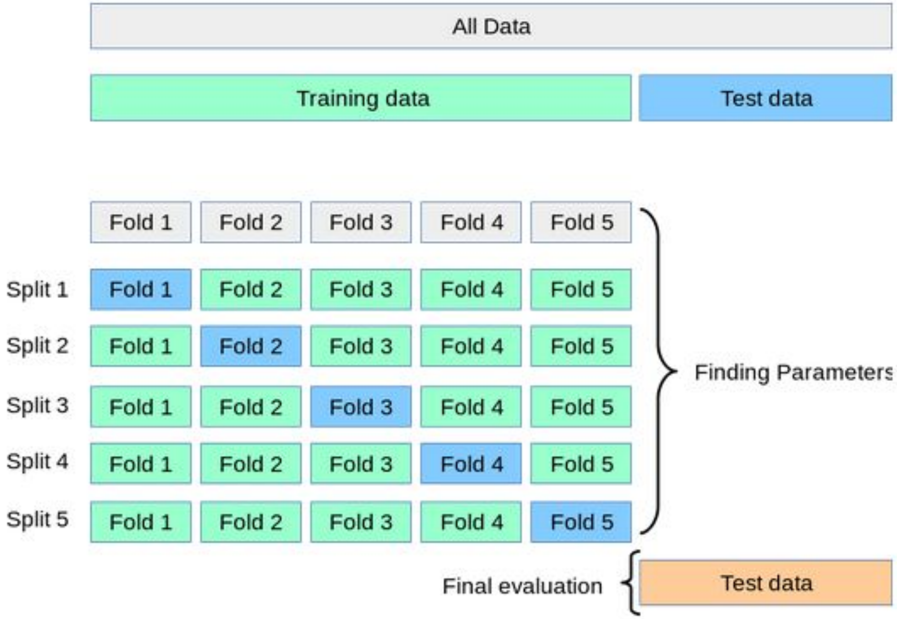


- **Cross Validation :**

1) **K-fold Cross Validation**
2) **Leave-One-Out Validation (LOOCV)**

Cont.

- Cross Validation

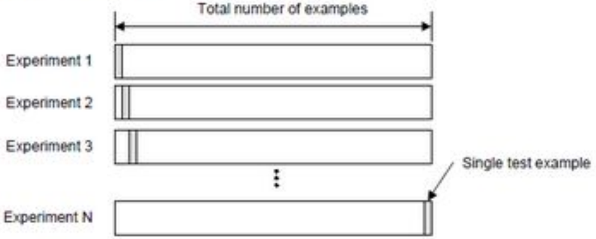


Cont.

- **Leave-One-Out Validation (LOOCV)**

Leave-one-out Cross Validation

- **Leave-one-out is the degenerate case of K-Fold Cross Validation, where K is chosen as the total number of examples**
  - For a dataset with N examples, perform N experiments
  - For each experiment use N-1 examples for training and the remaining example for testing



- **As usual, the true error is estimated as the average error rate on test examples**

$$E = \frac{1}{N}\sum_{i=1}^{N} E_i$$