PH-227
AI and Data Science

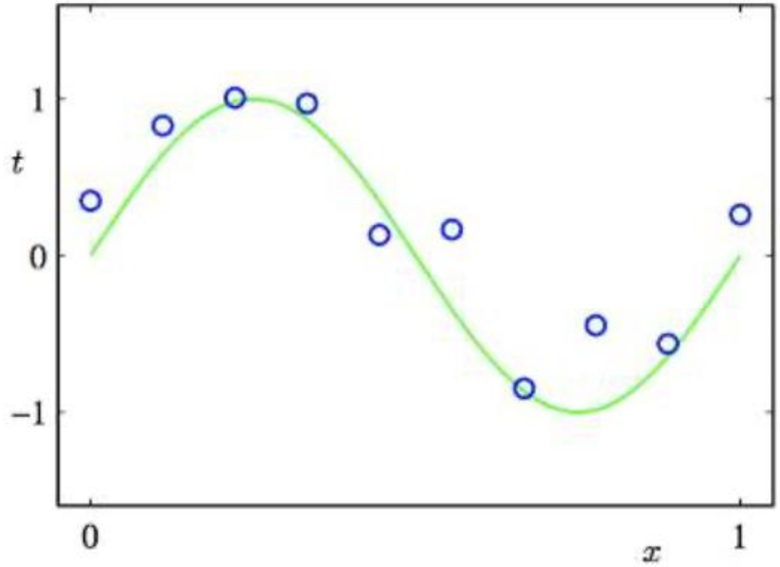
- Aftab Alam
- Email : aftab@iitb.ac.in
- Ext. : 5564 or 8564

TAs:

Yashowardhan, Divyansh, Matam, Peela, Piyush

A Quick Recap

Overfitting and Underfitting (Example)



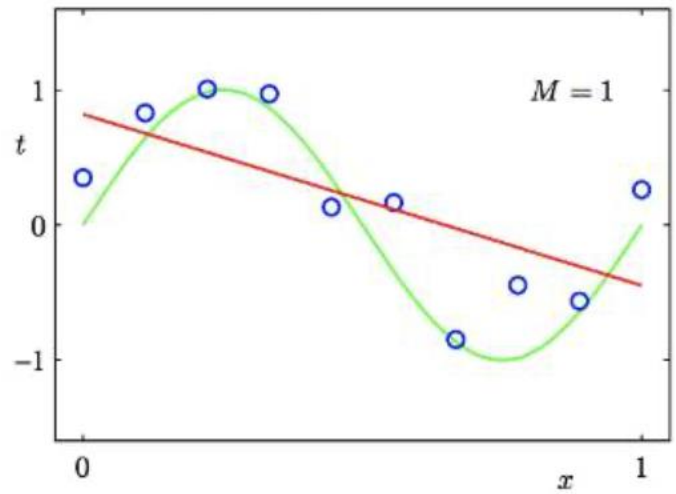
Blue circle: Raw data
Green line: Sin($2\pi x$)

- The goal is to find a model (or function) that fits the raw data well.
- But, we should be careful that the chosen model does not underfit or overfit !

Let us decide to fit a polynomial function of the following form,

$$y(x,w) = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + \dots\dots\dots\dots + w_M x^M$$

Overfitting and Underfitting (Example)

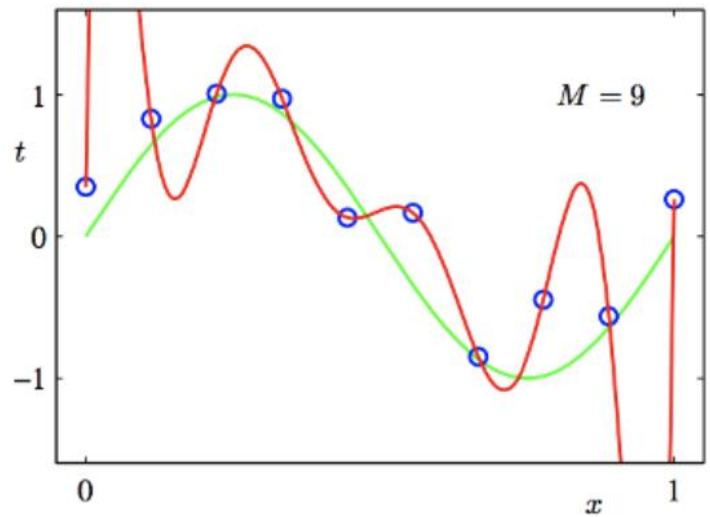


Blue circle: Raw data
Green line: Sin($2\pi x$)

$$y(x,w) = w_0 + w_1 x$$

- The model (with M-1) does not fit the data well, since it disagrees on many points in the training raw dataset.
- This is an example of underfitting because it is less expressive than the true function.

Overfitting and Underfitting (Example)

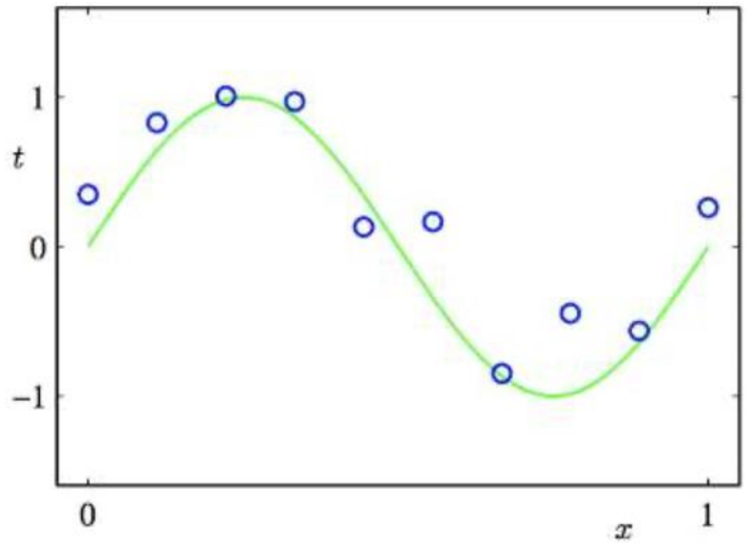


Blue circle: Raw data
Green line: $\text{Sin}(2\pi x)$

$$y(x,w) = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + \ldots\ldots\ldots\ldots + w_9 x^9$$

- Unlike the straight line (M=1), this function fits each of the points in the training set
- However, it does not approximate the function well at points not in the training set, indicating that it is overfitting.

Cross Validation (as a solution to overfitting)

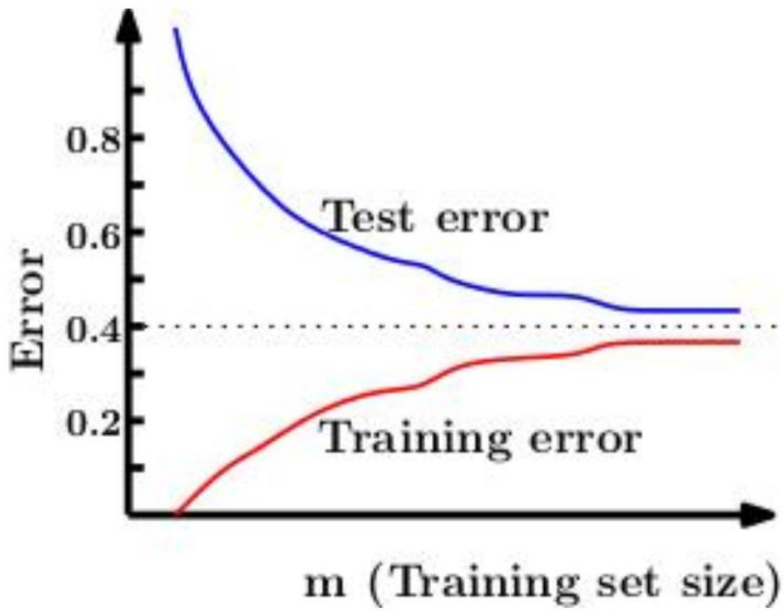


Blue circle: Raw data
Green line: Sin($2\pi x$)

- Cross validation detects overfitting by demonstrating how a single model+(parameter choices)+(training procedure) performs on several distinct test sets.
- It can be used to reduce overfitting by guiding model, parameter or training choices that lead to less overfit models, but it does not reduce overfitting for a single model+(parameter choices)+(training procedure) .
- The best way to learn something is to do it yourself.

Mean and Standard deviation

Training vs. Testing Error



Total training error: $\mathcal{E}_n(h; \quad) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}\left(h(x^{(i)}; \quad), y^{(i)}\right)$

Validation or Test error (n' new points): $\mathcal{E}(h) = \frac{1}{n'} \sum_{i=n+1}^{n+n'} \mathcal{L}\left(h(x^{(i)}), y^{(i)}\right)$

Ans: In the plot, the training error is increased with the training set size. The true error is around 0.4 which is quite high. This is the example of High bias, even if the testing error is decreasing.

Question on Learning

Q1: Suppose you are the marketing consultant of a leading e-commerce website. You have been given a task of making a system that recommends products to users based on their activity on Facebook.

You realize that user interests could be highly variable. Hence, you decide to do the following two tasks sequentially,

- **(Task1):** Cluster the users into communities of like-minded people and
- **(Task2):** Train separate models for each community to predict which product category (e.g., electronic gadgets, cosmetics, etc.) would be the most relevant to that community.

- What type of learning are Task1 and Task2?

Question on Learning

Q2: Which of the following tasks is NOT a suitable machine learning task(s) and why?

1) Finding the shortest path between a pair of nodes in a graph
2) Predicting if a stock price will rise or fall
3) Predicting the price of petroleum
4) Grouping mails as spams or non-spams

Ans: *Finding the shortest path between a pair of nodes in a graph is not a suitable machine-learning task because it falls under the category of graph algorithms and can be efficiently solved using algorithms like Dijkstra's algorithm. Machine learning is typically used for tasks that involve pattern recognition, prediction, or classification based on data. In this case, the task of finding the shortest path in a graph is better suited for algorithmic or graph theory-based approaches rather than machine learning.*

Question on Learning

Q3: One of the most common uses of Machine Learning today is in the domain of Robotics. Robotic tasks include a multitude of ML methods tailored towards navigation, robotic control and a number of other tasks. Robotic control includes controlling the actuators available to the robotic system. An example of this is control of a painting arm in automotive industries.

The robotic arm must be able to paint every corner in the automotive parts while minimizing the quantity of paint wasted in the process. Which of the following learning paradigms would you select for training such a robotic arm?

1) Supervised Learning
2) Unsupervised Learning
3) Combination of Supervised and Unsupervised Learning
4) Reinforcement Learning

Ans: *This kind of a learning problem warrants the use of Reinforcement Learning. We see that the robotic arm has to cover every corner, i.e. maximize the area covered and all the while minimizing the quantity of paint wasted in the process. One can design a primitive reward signal that takes into account the area covered and paint wasted (normalized to some extent) and use it to train a reinforcement learning agent.*

Further Discussion on Classification (predicting classes)

K-nearest neighbors (k-NN) algorithm

Example:

IMDb Rating	**Duration (mins)**	**Genre**
8.5 (Dune)	166	Action
7.2 (The Substance)	141	Horror
6.8 (Hit Man)	115	Comedy
5.8 (Scrambled)	100	Comedy

Task: Predict the genre of another movie “The Underdoggs” with IMDb rating 5.6 and duration 96 mins.

This is an example of a supervised learning

K-nearest neighbors (k-NN) algorithm

Step-1: Calculate the Euclidean distance between the new movie and each and each movie in the dataset

Co-ordinate of “The underdoggs” (5.6,96)

Distance from (8.5,166) $= \sqrt{(8.5-5.6)^2+(166-96)^2} \approx 70.06$ (Action)

Distance from (7.2,141) $= \sqrt{(7.2-5.6)^2+(141-96)^2} \approx 45.03$ (Horror)

Distance from (6.8,115) $= \sqrt{(6.8-5.6)^2+(115-96)^2} \approx 19.04$ (Comedy)

Distance from (5.8,100) $= \sqrt{(5.8-5.6)^2+(100-96)^2} \approx 4.00$ (Comedy)

Step-2: Select the K-nearest neighbors

For K=1, the smallest distance is chosen as the choice for genre

➔ “The underdogs” movie belong to the genre “Comedy”

Step-3: For K>1, majority voting classification is utlized

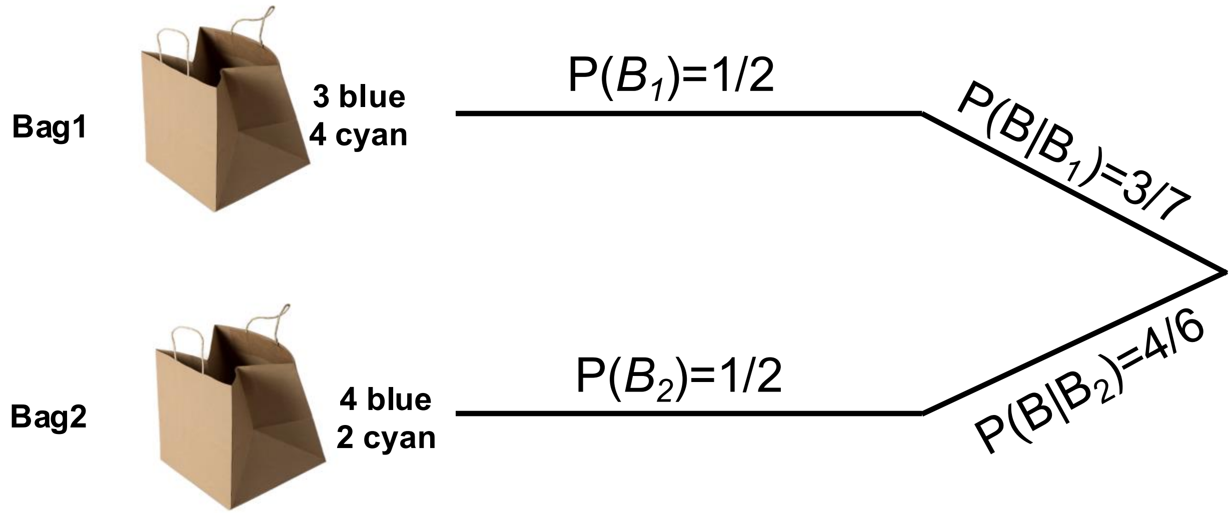
e.g. if we take K=3, the three smallest distance is chosen and their respective genres are screened and find the majority.

➔ Again, “The underdogs” movie belong to the genre “Comedy”

Naive Bayes' Classification

Let us first understand the *Bayes theorem*

Consider two identical bags (Bag1 & Bag2). Bag1 contains 3 blue and 4 cyan colored marbles, whiles Bag2 contains 4 blue and 2 cyan marbles. One ball is taken out at random from one of these bags and its blue. Find the probability that it is drawn from Bag1?



Total Probability ($P_{tot}(B)$)

$$(1/2)\times(3/7)+(1/2)*4/6$$
$$= 0.381$$

The task is to find the probability of picking a blue ball from Bag1

i.e. P(Bag1/Blue)

Bayes' Theorem

Conditional Probability

$$\mathrm{P(Y/X)} = \frac{P(X|Y) \times P(Y)}{P_{tot}(X)} \qquad (Y = which\ bag, X = color)$$

$$\mathrm{P}(\mathrm{Y}/X_1, X_2, X_3, \dots \dots X_n) = \frac{[P(X_1|Y) \times P(X_2|Y) \times P(X_3|Y) \dots \dots \dots .. \times P(X_n|Y)] \times P(Y)}{P_{tot}(X_1) \times P_{tot}(X_2) \times P_{tot}(X_3) \dots \dots \dots . P_{tot}(X_n)}$$

$$\mathrm{P}(B_1/\mathrm{B}) = \frac{P(B|B_1) \times P(B_1)}{P_{tot}(B)} = \frac{\left(\frac{2}{5}\right) \times \left(\frac{1}{2}\right)}{0.381} = 0.525$$

$$\mathrm{P(N/X)} = \frac{P(X|N) \times P(N)}{P_{tot}(X)}$$

$$\mathrm{P}(\mathrm{N}/X_1, X_2, X_3, \dots \dots X_n) = \frac{[P(X_1|N) \times P(X_2|N) \times P(X_3|N) \dots \dots \dots .. \times P(X_n|N)] \times P(N)}{P_{tot}(X_1) \times P_{tot}(X_2) \times P_{tot}(X_3) \dots \dots \dots . P_{tot}(X_n)}$$

Common Factor

Problem-1

Patient	Cold (Yes/No)	Cough (Yes/No)	Fever (Yes/No)
1	Yes	No	Yes
2	No	Yes	Yes
3	No	No	No
4	Yes	No	Yes
5	No	No	Yes
6	Yes	No	Yes
7	Yes	No	No
8	No	Yes	Yes
9	No	Yes	No

Step1: Prior Probability
P(Fever=Yes)=6/9
P(Fever=No)=3/9

Step2: Conditional Probability

	Fever	
	Yes	**No**
Cold	3/6	1/3
Cough	2/6	1/3

P(Yes | Cold,Cough)= (3/6)×(2/6)×(6/9) = 0.111
P(No | Cold,Cough)= (1/3)×(1/3)×(3/9) = 0.037

Q: If a person has both cold and cough, what is the possibility that he has Fever

P(Yes | Cold,Cough)=P(Cold | Yes) ×P(Cough | Yes) × P(Yes)

P(No | Cold,Cough)=P(Cold | No) ×P(Cough | No) × P(No)

Problem-2

Patient	Cold (Yes/No)	Cough (Yes/No)	Fever (Yes/No)
1	Yes	No	Yes
2	No	Yes	Yes
3	Yes	Yes	Yes
4	No	No	No
5	Yes	No	Yes
6	No	No	Yes
7	Yes	No	Yes
8	Yes	No	No
9	No	Yes	Yes
10	No	Yes	No

Step1: Prior Probability
P(Fever=Yes)=7/10
P(Fever=No)=3/10

Step2: Conditional Probability

	Fever: Yes	Fever: No
Cold	4/7	1/3
Cough	3/7	1/3

P(Yes | Cold,Cough)= (4/7)$\times$(3/7)$\times$(7/10) = 0.171
P(No | Cold,Cough)= (1/3)$\times$(1/3)$\times$(3/10) = 0.033

Q: If a person has both cold and cough, what is the possibility that he has Fever

P(Yes | Cold,Cough)=P(Cold | Yes) $\times$P(Cough | Yes) $\times$ P(Yes)

P(No | Cold,Cough)=P(Cold | No) $\times$P(Cough | No) $\times$ P(No)