CSCC11 Week 1 Notes

Introduction to Machine Learning:

- Machine Learning (ML) is a set of tools that allows computers to learn how to perform a task by giving examples of how it should be done.

- ML is usually broken into 2 steps:
  1. Training: A model is learned from a collection of training data.

  2. Application/Testing: The model is used to make decisions about some new test data.

- ML is all about "fitting" a function.

- Some fields in ML include:
  1. Supervised Learning: Here, the training data is labelled with correct answers. (Input/output) pairs
     Supervised learning has 2 main sub-topics:
     a) Regression: The target/output is a real-valued number.
     b) Classification: The target/output are discrete labels.

  2. Unsupervised Learning: Here, we are given a collection of unlabelled data, which we wish to analyze and discover patterns within. There are 3 main sub-topics:
     a) Density Estimation: Here, we estimate the parameters of a distribution that generated the data.
     b) Dimensionality Reduction: Reduce the dimension of high dimensional data (E.g. images)

c) Clustering: Here, we are grouping data with similar patterns together.

3. Reinforcement Learning: Here, an agent (I.e. a robot or controller) seeks to learn the optimal actions to take based on the state of the world, and hence the consequences of past actions.

4. Active Learning: Here, obtaining data is expensive and so an algo must determine which training data to acquire.

5. Meta Learning: Here, models that learned from tasks with large training sets or many tasks with moderate amounts of training data can be used to help constrain learning on related problems that have relatively small data sets.

   I.e. Machines observe how different ML approaches perform on a wide range of learning tasks and then learn from this experience to learn new tasks much faster.

   It is "learning to learn."

Math Formulas:

1. Mean Squared Error:
   - $L(y, \hat{y}) = \left(\sum_{n=1}^{N} (\hat{y}_n - y_n)^2 \right) \cdot \frac{1}{N}$

   y is the target and $\hat{y}$ is the prediction.
   $\hat{y} = f(x)$

- The Mean Squared Error (MSE) tells you how close a regression line is to a set of points. It does this by taking the differences or distances from the points to the regression line and squaring them. Note: The distances are the "errors". We square them to remove negative numbers and to give more weight to larger differences. The lower the MSE, the better the forecast.

2. Likelihood Function:
   - If $X_1, \dots X_n$ are an iid (independent and identically distributed) sample from a population with pdf or pmf $f(x|\theta_1, \dots, \theta_k)$, the likelihood is defined by

$$L(\theta|x) = L(\theta_1, \dots, \theta_k | X_1, \dots, X_N) = \prod_{i=1}^{N} f(X_i|\theta)$$

   - The likelihood function helps us find the best distribution of the data given a particular value of some feature or some situtation in the data.

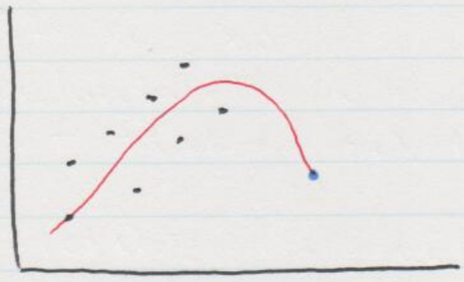
     I.e. Likelihood means to increase the chances of a particular situtation to occur by varying the characteristics of the dataset distribution.

   - A collection of random variables is iid if each r.v. has the same probability distribution as the others and all are mutually independent.
     E.g.
     1. Tossing a coin 10 times and counting the num of H.
     2. Rolling a die 5 times and counting how many times it lands on 6.

Overfitting and Underfitting:

- Often, data sets have outliers/noise that we don't want.
- Overfitting occurs when a model learns the data too well. In this case outliers/noises are also picked up. The problem is that these noises don't apply to new data and negatively impact the model's ability to generalize.

E.g. Suppose we have the data set below:



Here, our model is overfitted as it picked up the outlier.

Note: We don't want to fit perfectly as there could be noise.

- Underfitting refers to a model that can neither model the training data nor generalize to new data.

- This is how to choose your model:
  - Have a training set that's split into training and validation and a test set. Then:
    1. Train model on training.
    2. Test on validation.
    3. Repeat 1 and 2 until you find the model that performs best for the objective.
    4. Deploy such model on test set.

Validation is part of your training set that you set aside.
Test set is data not seen before.

Discriminative and Generative Models:

- **Discriminative models** learn the boundaries between classes or labels in a dataset.
- The goal of discriminative models is to separate one class from another.
- Discriminative models use $P(y|x)$

  Note: $P(y|x) = \dfrac{P(x|y) \cdot P(y)}{P(x)}$
- **Generative models** model the actual distribution of each class.
- They use $P(x, y)$
- E.g.

  One day, a father takes his 2 kids, A and B, to a zoo to see a lion and an elephant only. After seeing both animals, the father takes out a picture and asks both children if the picture is of a lion or elephant.

  kid A draws a picture of a lion and elephant on a piece of paper based on what he saw in the zoo. Then, after comparing his pictures with his father's, he says "A lion".

  kid B only knows the differences based on diff properties of the 2 animals and also says "A lion".

  kid A is an example of a generative model while kid B is an example of a discriminative model.