**Natural Language Processing (NLP)**

**Session 2**

1. **Text Classification**:
   1. *Supervised*:
      1. *Input*: a document, a fixed set of classes and a training set of labelled data
      2. *Output*: a learned classifier / model
         1. Generative: build a model by aggregating the features of a class e.g., Naïve Bayes.
         2. Discriminative: build a model by separating the features of a class e.g., logistic regression.
2. **Naïve Bayes**: Based on Bayes Rule (ML – Notes, Week 2, Session 2, Page 2). If we make very naive assumptions about the generative model for each label, we can find a rough approximation of the generative model for each class, and then proceed with the Bayesian classification.
   1. To prevent underflow of numerical probabilities, summing up of logarithmic probabilities can be implemented.
   2. Advantages:
      1. Works quickly and can save a lot of time.
      2. Suitable for solving multi-class prediction problems.
      3. If its assumption of the independence of features holds true, it can perform better than other models and requires much fewer training data.
   3. Disadvantages:
      1. Naive Bayes assumes that all predictors (or features) are independent, rarely happening in real life. This limits the applicability of this algorithm in real-world use cases.
      2. This algorithm faces the ‘zero-frequency problem’ where it assigns zero probability to a categorical variable whose category in the test data set wasn’t available in the training dataset. It would be best if you used a smoothing technique to overcome this issue.
      3. Its estimations can be wrong in some cases, so you shouldn’t take its probability outputs very seriously.
3. ***Maximum a posteriori* (MAP) hypothesis for classification**:
   1. Density estimation is the problem of estimating the probability distribution for a sample of observations from a problem domain.
   2. MAP estimation is a probabilistic framework for solving the problem of density estimation.
   3. MAP involves calculating a conditional probability of observing the data given a model weighted by a prior probability or belief about the model.
   4. MAP provides an alternate probability framework to maximum likelihood estimation for machine learning.
4. **Smoothing**: Zero probabilities (noise) cannot be conditioned away, however for training purposes, the noise can be problematic i.e., infinity values, etc. Data smoothing uses an algorithm to remove noise from a data set, allowing important patterns to stand out.