**Chapter 2**

2.1 Introduction to regression models

* The standard model for regression is , where is the ***reducible* error** (the error in prediction can be potentially reduced by choosing a different model) and is the ***irreducible error*** (it originates from e.g. measurement errors or inherent stochasticity and cannot be reduced by changing the model).
* A good will allow us to make predictions of the value of at new points of .
* The ideal function will give the expectation value of at :

This ideal function is called the ***regression function***. This ideal function is optimal is the optimal predictor of in the sense that it will minimize the ***sum of squared errors***.

* Typically, there will be very few, if any, data points that correspond to a certain value of (for example, there may be no values for the output data at exactly ). In that case the expectation value cannot be determined.
* In this case, is estimated based on the points in the ***neighbourhood*** of :

Where is some neighbourhood around . This is called ***Nearest Neighbour or Local averaging***.

2.2 Dimensionality and Structured Models

* Nearest neighbour averaging becomes more problematic as the number of dimensions increases: For a high number of dimensions, the width of the neighbourhood for nearest neighbour averaging needs to increase to capture the same number of data points.
* As a result, nearest neighbour averaging becomes less local for as the number of dimensions increases. This is called the ***Curse of Dimensionality***.
* The curse of dimensionality can be circumvented by using ***structural models*** that do not depend local properties and nearest neighbour averaging. An example of a structural model is a linear model for which each of the parameters can be found by fitting it to all datapoints. As a result, these linear models do not depend on any local properties or nearest neighbour averaging.
* Structural models that are less ***flexible*** (for example, linear vs. quadratic models) typically have lower ***interpretability***, although they may provide a better fit to the data.
* It is important to choose a model that is not too flexible to prevent ***overfitting*** of the data.

2.3 Model Selection and Bias-Variance Tradeoff

* To see how well a model performs, it can be fitted to a set of ***training data***:

However, this can be ***biased towards more overfit models***. Instead, the value should be evaluated on a set of ***test data***:

* The performance of models with different flexibilities can be assessed by plotting against model flexibility. The value of will typically have a minimum value for a certain flexibility of the model, while will keep decreasing as the model becomes more flexible.
* The choice for a certain flexibility of the model is subject to a ***Bias-Variance Tradeoff***: the variance in predictions of a model will increase as the model becomes more flexible (the model will become more sensitive to the peculiar characteristics of the training dataset). In return, the bias will decrease as the model becomes more flexible.

2.4 Classification

* In classification, the response variable is ***qualitative*** and the aim is to build a classifier that can assign a class label from the set of labels for future observations of .
* An ideal classifier is the ***Bayes Optimal Classifier*** that minimizes the probability of misclassification.
* If there are elements in the set of class labels then the ***Conditional Class Probabilities*** are:

The Bayes Optimal Classifier at is given by:

In other words, the Bayes Optimal Classifier assigns the observation to the ***most class with the highest probability***.

* In classification, nearest neighbour averaging can be used as in logistic regression (take the conditional probabilities in the neighbourhood of ). In that case, the curse of dimensionality still applies. However, the curse of dimensionality has less impact on than on .
* The performance of the classifier is typically measured using the ***Misclassification Rate***:

In words: ***the error is the average number of mistakes***. This error is the smallest for the Bayes Classifier.

* ***Support Vector Machines*** build structured models for **.**