Saarland University

The Elements of Stastical Learning

Assignement 1

Due Date: 08.11.2017

Thibault SCHOWING

November 6, 2017

Problem 1

Statistical learning is the process of learning from data. The majority of statistical learning can be divided into two categories: **Supervised Learning** and **Unsupervised learning**. In Supervised Learning, you have **inputs** as well as the corresponding **outcomes** that serve to direct the learning process. The goal is to estimate some unknown function f, which serves as the information about the relationship between inputs and outputs. In Unsupervised Learning we observe only the predictors; the responses are not available. The goal of unsupervised learning is to discover something about the data, for example how it groups together. For this reason it is often referred to as Clustering.

Supervised Learning is used for **inference** and **prediction**. In Prediction the goal is to estimate some unknown function f so as to accurately predict some output given a new input. This can be done for **quantitative** values, which is known as **regression** (on continuous values), or for **qualitative** (or categorical) values, which is referred to as **classification**.

When talking about supervised learning it's also important to discuss **training data** and **test data**. The training data is used, as the name suggests, to train a statistical model. Test data are data that have not been given to the model during training and are used to test the performance of that model.

The methods used to to estimate f can be categorized as **parametric** or **non-parametric**. Parametric methods make the assumption that the unknown function is linear, and use linear methods to estimate it. Non-parametric methods make no assumptions about the underlying form of f, and thus can use methods that are much more coplicated and have many more degrees of freedom.

Problem 2

Show that:

$$E(Y) = argmin_c E[(Y - c)^2]$$

We are looking to proof that the value of c for which $E[(Y-c)^2]$ attains its minimum is $E(Y)$.

Development:

$(Y-c)^2 = Y^2 - 2cY + c^2$

$E[(Y-c)^2] = E[Y^2 - 2cY + c^2] = E[Y^2] - 2cE[Y] + c^2$

We assume here that c is a constant, and so the derivative is:

$\frac{d}{dc}E[Y^2] - 2cE[Y] + c^2 = -2E(Y) + 2c$

We equal it to zero and obtain:

$2c = 2E(Y)$ and so $c = E(Y)$ which show that the first statement is true.

Problem 3

Prove the bias-variance trade-off with irreducible error:

$E[(y_0 - \hat{f}(x_0) - E(\hat{f}(x_0)))^2] + [E(\hat{f}(x_0) - f(x_0)]^2 + Var(\epsilon) =$

$Var(\hat{f}(x_0)) + [Bias(\hat{f}(x_0))] + Var(\epsilon)$

Frist we note $E[(y_0 - \hat{f}(x_0))^2] = E[(y_0 + \epsilon - \hat{f}(x_0))^2]$ Next, we expand into $E[y_0^2 + \hat{f}(x_0)^2 - 2y_0\hat{f}(x_0)]$ Assuming that y_0 is deterministic, we can simplify to $E[\hat{f}(x_0)^2] + y_0^2 - 2y_0E[\hat{f}(x_0)]$