Saarland University

The Elements of Stastical Learning

Assignement 1

Due Date: 08.11.2017

Thibault SCHOWING

October 26, 2017

Problem 1

Statistical learning is divided in two main groups: Supervised and Unsupervised learning.

Supervised learning will try to find a prediction function from data. In those data we have one or more input variables, also called inputs, features, predictors or independent variables and usually one output variable, called outcome, response or dependent variables. The prediction can be whether regression, if the response variable is a quantitative variable or classification, if the response variable is a qualitative (or categorical) variable. Estimating those output variables means making a prediction.

In order to test the model and to know its performance, we usually separate the model in two parts: Training data and test data. Training data will be used to create and train the model. Test data will have the role of new data, not used in the training phase, to test the efficiency of the model. Repeating the train-test operation with different groups of train-test data is called cross-validation.

Once we have a prediction function, we might want to understand the relations between the variables in order to find unknown relations in the data. This is inference. Often, if a prediction model is very accurate, it will be harder to interpret.

To estimate the prediction function, we can use parametric or non-parametric methods. With the parametric method, we take a classical functional form (like linear, quadratic. . .) and fit the function my adjusting the function's parameters. For the non-parametric method, we have more freedom in choosing the form but we have to choose a lot of parameters that requires a lot of observation and there is a risk of overfitting (modelling the noise).

Unsupervised learning is used when there is no output data. The goal is to find different groups (or clusters) in the data by finding relationships between the variables or the observations.

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.

What is the practical benefit of this equation when using the squared error as error metric?

??????????????????????????????????

Problem 3

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

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

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

One of the properties of variance is that $Var(X) = E[X^2] - (E[X]^2)$

We replace X with $y_0 - \hat{f}(x_0)$ and so obtain that $Var(y_0 - \hat{f}(x_0)) = E[(y_0 - \hat{f}(x_0))^2] - (E[y_0 - \hat{f}(x_0)]^2)$.

In this equation we recognize that $E[y_0 - \hat{f}(x_0)]^2 = MSE$

And we also have $Bias = E[y_0 - \hat{f}(x_0)]$