**DECISION FUNCTION**

A decision function is a function which takes a dataset as input and gives a decision as output. What the decision can be depends on the problem at hand. Examples include:

Estimation problems: the "decision" is the estimate.

Hypothesis testing problems: the decision is to reject or not reject the null hypothesis.

Classification problems: the decision is to classify a new observation (or observations) into a category.

Model selection problems: the decision is to choose one of the candidate models.

Typically, there are an infinite number of decision functions available for a problem. If we for instance are interested in estimating the average height of Swedish males based on ten observations x=(x1,x2,…,xn), we can use any of the following decision functions d(x):

The sample mean: d(x)=110∑10i=1xi.

The median of the sample: d(x)=median(x)

The geometric mean of the sample: d(x)=x1⋯x10−−−−−−−√10

The function that always returns 1: d(x)=1, regardless of the value of x. Silly, yes, but it is nevertheless a valid decision function.

How then can we determine which of these decision functions to use? One way is to use a loss function, which describes the loss (or cost) associated with all possible decisions. Different decision functions will tend to lead to different types of mistakes. The loss function tells us which type of mistakes we should be more concerned about. The best decision function is the function that yields the lowest expected loss. What is meant by expected loss depends on the setting (in particular, whether we are talking about frequentist or Bayesian statistics).

In summary:

Decision functions are used to make decisions based on data.

Loss functions are used to determine which decision function to use.

**1.2.6.1.3. Parameters of the RBF Kernel**

When training an SVM with the *Radial Basis Function* (RBF) kernel, two parameters must be considered: C and gamma. The parameter C, common to all SVM kernels, trades off misclassification of training examples against simplicity of the decision surface. A low C makes the decision surface smooth, while a high C aims at classifying all training examples correctly. gamma defines how much influence a single training example has. The larger gamma is, the closer other examples must be to be affected.

Proper choice of C and gamma is critical to the SVM’s performance. One is advised to use GridSearchCV with C and gamma spaced exponentially far apart to choose good values.

C=0.8 gamma=0.005