Error functions - classification

Kristian Wichmann

February 27, 2017

1 Continous-valued predictors

Most predictors for classification will be based on a continuous-valued score for each of the two outcomes. The prediction is then made based on these values (usually, the one with the highest score is chosen).

Given the values from such a predictor, we would like to have some quantitative measure of how good the prediction is.

1.1 Terminology

Consider classification between n cases. If the true classification for a sample x is k, then we define the n-dimensional vector t by $t_i = \delta_{ik}$. Let the values made by the predictor be the vector h, also a n-dimensional vector.

2 Error functions

2.1 Squared error

In this scheme we use the same error as used for regression - the squared error:

$$\frac{1}{2}||t - h||^2 = \frac{1}{2}(t - h)^T(t - h) = \frac{1}{2}\sum_{i=1}^{n}(t_i - h_i)^2 \qquad (2.1)$$

2.2 Cross-entropy error

This error is based on the log-likelihood of a Bernoulli experiment:

$$\sum_{i=1}^{n}\left[-t_i \log(h_i) - (1 - t_i)\log(1 - h_i)\right] \qquad (2.2)$$

3 Forms of the hypothesis h

The hypothesis h is in general the function of some input feature vector x, so $h = h(x)$. Here we will consider only hypotheses which depends on the vector x weighted by a vector of weights w_i depending on the outcome i. In other words, we only look at $h_i(x) = h_i(w_i^T x)$.

3.1 Linear hypothesis

Here, the hypothesis is simply equal to $w_i^T x$, so that:

$$h_i(x) = w_i^T x \tag{3.1}$$

3.2 Logistic hypothesis

Here, the hypothesis is the logistic function of $w_i^T x$. The logistic function is:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{3.2}$$

And so, the hypothesis is:

$$h_i(x) = \sigma(w_i^T x) \tag{3.3}$$

3.3 Softmax hypothesis

Here, the softmax function is used to normalize the hypothesis, so that they sum to 1. The softmax is defined by:

$$s_i(z) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \tag{3.4}$$

And the hypothesis:

$$h_i(x) = s_i(w^T x) = \frac{e^{w_i^T x}}{N(w, x)} \tag{3.5}$$

Here, w is a matrix in which the columns are the weights w_i, and $N(w, x)$ is the normalization constant $\sum_{j=1}^{n} e^{z_j}$.

4 Combining error and hypothesis types

4.1 Squared error with logistic hypothesis

Combining equations 2.1 and 3.3 we get the error function:

$$\frac{1}{2} \sum_{i=1}^{n} (t_i - \sigma(w_i^T x))^2 \tag{4.1}$$

4.2 Cross-entropy error with logistic hypothesis

Combining equations 2.2 and 3.3 we arrive at the error function:

$$\sum_{i=1}^{n} \left[-t_i \log(\sigma(w_i^T x)) - (1 - t_i) \log(1 - \sigma(w_i^T x)) \right] \tag{4.2}$$

4.3 Cross-entropy error with softmax hypothesis

Combining equations 2.2 and 3.5 we arrive at the error function:

$$\sum_{i=1}^{n} \left[-t_i \log\left(\frac{e^{w_i^T x}}{N(w,x)} \right) - (1 - t_i) \log\left(1 - \frac{e^{w_i^T x}}{N(w,x)} \right) \right] \tag{4.3}$$