Recognition Labs – Maximal Likelihood Parameter Estimation

Jan Šochman, Vojtch Franc, Petr Doubek

October 23, 2009

Let $p(x|y,\theta)$ be probability density with unknown parameter θ. In the following text, we will consider only one class (we will estimate the probability density of each class separately, i.e. we assume they are independent) and use simplified notation $p(x|\theta)$. Let's have a training set

$$T = \{x_1, \dots, x_n\}.$$

Members of set T (called measurements or samples) are obtained by random selection from distribution $p(x|\theta)$. The measurements have to be independent. Our goal is to find maximal likely estimate of parameter θ given the training set T.

The quality of estimate θ with respect to the training set T is measured by conditional probability $P(T|\theta)$, called *likelihood*. Thanks to independency of training samples the likelihood can be evaluated as

$$P(T|\theta) = \prod_{i=1}^{n} p(x_i|\theta). \tag{1}$$

Sometimes also the logarihmic form is used

$$L(T|\theta) = \sum_{i=1}^{n} \log p(x_i|\theta). \tag{2}$$

We are searching θ^* which maximizes the likelihood

$$\theta^* = \operatorname*{argmax}_{\theta} P(T|\theta) = \operatorname*{argmax}_{\theta} \prod_{i=1}^{n} p(x_i|\theta). \tag{3}$$

1 Finding Maximum Likelihood Estimate

A transformation by logarithm keeps the point where the function has its maximum, therefore we can transform equation (3)

$$\theta^* = \operatorname*{argmax}_{\theta} \sum_{i=1}^{n} \log p(x_i|\theta) \tag{4}$$

The term to be maximized is

$$L(T, \theta) = \sum_{i=1}^{n} \log p(x_i|\theta). \tag{5}$$

For convex L, the maximum is found at point where its derivation equals zero

$$\frac{\partial L(T, \theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \log p(x_i|\theta)}{\partial \theta} = 0. \tag{6}$$

2 Example – One-Dimensional Normal Distribution

Let us look at the example of one-dimensional normal distribution

$$p(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{7}$$

Assume the standard deviation σ known and training set $T = \{x_1, \ldots, x_n\}$ given. We search for μ such that

$$\sum_{i=1}^{n} \frac{\partial \log p(x_i|\mu)}{\partial \mu} = 0. \tag{8}$$

By simplification and derivation we obtain

$$\sum_{i=1}^{n} \frac{(x_i - \mu^*)}{\sigma^2} = 0. \tag{9}$$

This term is zero when

$$\sum_{i=1}^{n} (x_i - \mu^*) = 0. \tag{10}$$

Rearranging the term we arrive at result

$$\mu^* = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{11}$$

We found the maximum likelihood estimate of mean of normal probability distribution. Similarly, we can fix the mean μ and solve for standard deviation, obtaining the formula

$$\sigma^* = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2} \tag{12}$$