Homework 2 – Maximum likelihood estimation

Machine Learning / Greg Hamerly

January 29, 2007

This homework is due in class on Friday, February 2, 2007.

For this homework, you will show that the maximum likelihood estimate for the linear model with additive Gaussian noise leads to the same result as the minimum RSS (Residual Sum of Squared errors) estimation. For this homework, please use LaTeXto formulate your solutions in a nice-looking format, and describe your work as you go; do not just put equations on paper.

Recall that our statistical model for linear regression is $Y = f(X) = \beta^T X + \varepsilon$ where we assume that $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ (the normal, or Gaussian, distribution) and the errors are i.i.d. Recall also that if $Z \sim \mathcal{N}(\mu, \sigma^2)$, then

$$Pr(Z = z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-(z - \mu)^2/\sigma^2\right)$$

and therefore we can express the distribution of the output of our problem as $Y \sim \mathcal{N}(f(X), \sigma^2)$ and the probability of one example $\{x_i, y_i\}$ from our problem is

$$Pr(Y = y_i|x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-(y_i - f(x_i))^2/\sigma^2\right)$$

Let's consider the simple 1-dimensional input regression problem with one coefficient to estimate – the slope. Then our particular regression model is $f(X) = \beta X$, where X is a scalar and not a vector. Use the above facts and this model to do the following:

1. Set up the likelihood function, $L(\beta)$, and express it in as much detail as you can.
2. Use the likelihood function to get the log-likelihood function $\mathcal{L}(\beta)$, and again simplify and use as much detail as you can.
3. Take the derivative of $\mathcal{L}$ with respect to β, and use the derivative to find the maximum likelihood estimate of β.
4. Explain in your own words what this approach – maximizing the likelihood function – is really doing.

Now consider the problem from the least-squares perspective, where we are trying to minimize

$$RSS(\beta) = \sum_{i=1}^{n} (y_i - \beta x_i)^2$$

Normally, the estimate of β is $\hat{\beta} = (X^T X)^{-1} X^T y$, but here our problem is simplified due to having only one parameter to estimate. Starting from this definition, do the following:

1. Take the derivative of $RSS(\beta)$ with respect to β.
2. Find the least-squares estimate of β by setting the derivative you just took to zero and solving for β.
3. Show that the estimate you got ($\hat{\beta}$) is equivalent to the maximum likelihood estimate. Also, show that it is equivalent to $(X^T X)^{-1} X^T y$. Explain these results.