Machine Learning 4/M

Maximum Likelihood Laboratory

Aims

- To implement the maximum likelihood estimator of the parameters of a linear model
- To plot predictions and their variance

Tasks

- Download the Olympic data (again)
- Implement the maximum likelihood estimator for the parameters $\mathbf{w}$ and σ^2 of the linear model
- Note that $\mathbf{w}$ should be identical to the value from minimising the loss
- The relevant equations are:

$$\mathbf{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{t} \ \ \text{(a vector with one value per parameter)}$$

$$\sigma^2 = \frac{1}{N}(\mathbf{t} - \mathbf{X}\mathbf{w})^T(\mathbf{t} - \mathbf{X}\mathbf{w}) \ \ \text{(a scalar)}$$

- Plot the training data, the predictive mean (i.e. $\mathbf{X}_{test}\mathbf{w}$, the polynomial function)
- On top of your previous plot add dashed lines to show $\pm\sigma$, i.e. a line at $\mathbf{X}_{test}\mathbf{w} + \sigma$ and one at $\mathbf{X}_{test}\mathbf{w} - \sigma$
- Plot the predictive density for the 2016 Olympics (your x axis will be winning time, t, and your y axis p(t)). I.e. a Gaussian pdf with mean $\mathbf{w}^T\mathbf{x}_{2016}$ and variance σ^2

Additional task (non-programming)

If you want a better understanding of the idea of maximum likelihood, this is a useful exercise to do.

Assume you have observed 10 numbers and you make the assumption that these numbers came, independently, from a Gaussian distribution (i.e. they came from a random number generator that used a Gaussian curve for its density).

You would like to *fit* a Gaussian to this data using maximum likeliood. Derive the maximum likelihood estimate of the mean of the Gaussian.

Hints

Because you are assuming that the data come independently from the same Gaussian, the likelihood is the product of Gaussian pdfs evaluated at each of the observed values:

$$L = \prod_{n=1}^{N} \mathcal{N}(x_n | \mu, \sigma^2)$$

where μ is the mean and σ^2 the variance. This is equal to:

$$L = \prod_{n=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x_n - \mu)^2\right\}$$

and to derive the desired estimator, you should log this expression and then differentiate with respect to μ, set to zero, and solve.