Machine Learning I WS2016/17

Lecturer: Prof. Dr. Bethge

2. Maximum Likelihood Learning and Bayesian Inference

Task 1: Linear regression (MLE):

Download the dataset `http://socr.ucla.edu/docs/resources/SOCR_Data/SOCR_Data_Dinov_020108_HeightsWeights.html`. Let $x_1, \dots, x_n$ denote the height and $y_1, \dots, y_n$ the weight. Split the dataset into a training set with 20.000 samples and a test set with 5.000 samples. Assume that the data is i.i.d. dan can be described by a Gaussian likelihood

$$p(x, y|\mu_x, \sigma_x^2, a, b, \sigma_n^2) = p(y|x, a, b, \sigma^2)p(x|\mu_x, \sigma_x^2) = \mathcal{N}(y|ax + b, \sigma_n^2)\mathcal{N}(x|\mu_x, \sigma_x^2).$$

Determine how to compute the parameters for which the data likelihood takes a maximum. Compute the MLE for $n = 100, 200, \dots, 20.000$ and plot the log-likelihood on the training set and on the test set as a function of n in one plot.

Task 2: Linear regression (Bayesian inference):

Use the same training and test sets and the Gaussian likelihood as in task 1. In addition, choose a Gaussian prior distribution over the parameters $\mu_x, \sigma_x^2, a, b, \sigma_n^2$ and

a) motivate your choice for the specific prior you picked.

b) Compute the posterior distribution over $\mu_x, \sigma_x^2, a, b, \sigma_n^2$.

c) Compute the MLE for $n = 100, 200, \dots, 20.000$ and plot the log-likelihood on the training set and on the test set as a function of n in the same plot you prepared for task 1.

d) Compute the information gain $I[\mu_x, \sigma_x^2, a, b, \sigma_n^2 : \mathcal{D}_n] = h[\mu_x, \sigma_x^2, a, b, \sigma_n^2] - h[\mu_x, \sigma_x^2, a, b, \sigma_n^2|\mathcal{D}_n]$ as a function of $n = 100, 200, \dots, 20.000$ where $\mathcal{D}_n$ stands for the data used to compute the posterior.

Task 3: Conjugate prior:

Use the shape-rate parametrization

$$\rho(\mu) = g(\mu|\alpha, \beta) = \frac{\mu^{\alpha-1}e^{-\beta\mu}}{\beta^{-\alpha}\Gamma(\alpha)}$$

of the Gamma distribution to show that it can serve as a conjugate prior for the Poisson distribution $\mathcal{P}_{Poisson}(k|\mu)$. That is, you have to show that $\rho(\mu|k) = g(\mu|\alpha_{new}(\alpha, \beta, k), \beta_{new}(\alpha, \beta, k))$.

Task 4: Read (and understand) the following text:

`http://www.cs.cmu.edu/%7Etom/mlbook/NBayesLogReg.pdf`