N. Bertschinger
M. Kaschube

Machine Learning II
3. Exercise Sheet
Discussed on Wed, May 9, 14:15

Exercise 1. *Suppose you have observed N samples $k_1, \ldots, k_N$ drawn from a Poisson distribution. Compute the maximum likelihood estimator for the rate of the data, i.e.*

$$\hat{\lambda} = \text{argmax}_\lambda \prod_{n=1}^{N} p(k_n; \lambda)$$

where $p(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!}$. **2 points**

Exercise 2. *Consider the linear regression model, i.e.*

$$p(\mathbf{t}|\mathbf{X}, \theta) = \prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^T \mathbf{x}_n, \sigma_\epsilon)$$

with design matrix $\mathbf{X}$ and target outputs $\mathbf{t}$.

Now, compute the Bayesian posterior $p(\mathbf{w}|\mathbf{X}, \mathbf{t}, \sigma_\epsilon^2)$, assuming that the noise variance σ_ϵ^2 is known and using a (multi-variate) Gaussian prior

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \sigma_0^2 \mathbf{I}).$$

3 points

Exercise 3. *Show that the evidence of the Bayesian linear regression model is given by*

$$\ln p(\mathbf{t}|\mathbf{X}) = \frac{D}{2} \ln \tau_0 + \frac{N}{2} \ln \tau_\epsilon - \frac{N}{2} \ln 2\pi - \frac{\tau_\epsilon}{2} (\mathbf{X}\mu_N - \mathbf{t})^T (\mathbf{X}\mu_N - \mathbf{t}) - \frac{\tau_0}{2} \mu_N^T \mu_N + \frac{1}{2} \ln |\mathbf{\Sigma}_N|$$

where N denotes the number of data points, D the number of inputs and the posterior parameters are given by

$$\begin{aligned} \mu_{\mathbf{N}} &= \tau_\epsilon \mathbf{\Sigma}_N \mathbf{X}^T \mathbf{t} \\ \mathbf{\Sigma}_N &= \left(\tau_0 \mathbf{I} + \tau_\epsilon \mathbf{X}^T \mathbf{X} \right)^{-1} \end{aligned}$$

2 points

Exercise 4. *Fit a linear model to the data in* data_simpson.csv *that can be downloaded from the course website.*

- *What if I told you that x denotes annual school teacher's salary (in thousands of US dollars) and y is the average total SAT score?*
- *Interpret your fit in the light of this background information.*

N. Bertschinger
M. Kaschube

2 points

Exercise 5. *Replicate figure 3.9 of PRML illustrating Bayesian linear regression:*

- *Data sets in this example were generated by drawing* $N = 1, 2, 4$ *and* 25 *points* (x_n, t_n) *with*

$$\begin{aligned} x_n &\sim Uniform(0,1) \\ t_n &= \sin(2\pi x) + 0.3\epsilon \end{aligned}$$

  where $\epsilon \sim \mathcal{N}(0,1)$.

- *A linear model with* 9 *radial basis functions equally spaced between* 0 *and* 1 *was then fitted to the data:*
  - *Assume that the noise precision is known, i.e.* $\tau_\epsilon = \frac{1}{0.3^2}$, *and use a prior precision of* $\tau_0 = 1$.
  - *Find a suitable width for the the basis functions!*

3 points