N. Bertschinger
M. Kaschube
V. Ramesh

Machine Learning I
6./7. Exercise Sheet
Due on Friday, Dec 4, 14:15

Exercise 1. *Reproduce the plots shown in Figure 1 (from the book by CM Bishop (Springer Series), p.155), using a slightly more complicated model. Generate your own synthetic data from the function*

$$f(x, \mathbf{a}) = a_0 + a_1 x + a_2 x^2$$

with parameter values $a_0 = -0.3$, $a_1 = 0.5$, $a_2 = 0.4$ *by first choosing values of* x_n *from the uniform distribution* $U(x|-1, 1)$, *then evaluating* $f(x_n, \mathbf{a})$, *and finally adding Gaussian noise with standard deviation of s=0.2 to obtain the target values* t_n. *The goal is to recover the values of* a_0, a_1 *and* a_2 *from such data, and to explore the dependence on the size of the data set. To achieve this, assume a model in which individual data points are generated by*

$$p(t|x, \mathbf{w}) = \mathcal{N}(t|y(x, \mathbf{w}), s^2) \ ,$$

where $y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2$ *with weights* $\mathbf{w}$ *to be estimated and a fixed standard deviation* $s = 0.2$, *i.e. assumed to be known. The likelihood is then given by*

$$p(\mathbf{t}|X, \mathbf{w}) = \prod_{n=1}^{N} \mathcal{N}(t_n|y(x_n, \mathbf{w}), s^2) \ .$$

Finally, assume a Gaussian distributed prior $p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha)$ *with* $\alpha = 2$. *Generate two plots analog to those shown in Figure 1, one for* (w_0,w_1) *and one for* (w_1,w_2). *Describe and interpret these plots thoroughly.*

5 points

Exercise 2. *Let* $p(x|\mu)$ *be a univariate Gaussian* $\mathcal{N}(\mu, \sigma^2)$ *with unknown parameter mean, which is also assumed to follow a Gaussian* $\mathcal{N}(\mu_0, \sigma_0^2)$. *From the theory exposed before we have*

$$p(\mu|X) = \frac{p(X|\mu)p(\mu)}{p(X)} = \frac{1}{\alpha} \prod_{k=1}^{N} p(x_k|\mu)p(\mu)$$

where for a given training data set, X, $p(X)$ *is a constant denoted as* α. *Write down the explicit expression for* $p(\mu|X)$.

1 points

N. Bertschinger
M. Kaschube
V. Ramesh

Machine Learning I
6./7. Exercise Sheet
Due on Friday, Dec 4, 14:15

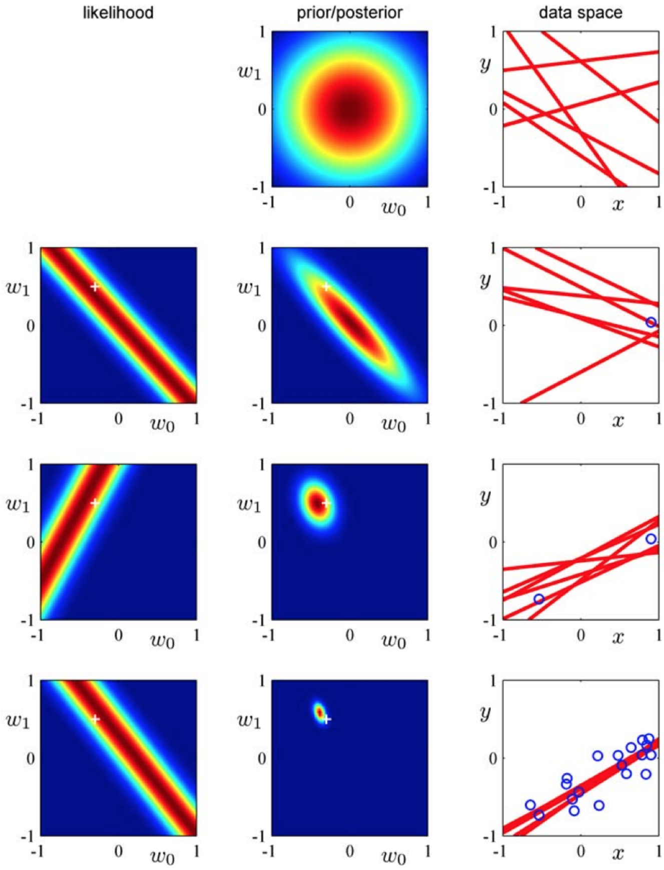

Figure 1: Illustration of sequential Bayesian learning for a linear model.

Exercise 3. *Show that, given a number of samples, N, the posterior $p(\mu|X)$ turns out to be also Gaussian, that is*

$$p(\mu|X) = \frac{1}{\sigma_N\sqrt{2\pi}} \exp\left(-\frac{(\mu-\mu_N)^2}{2\sigma_N^2}\right)$$

with mean value

$$\mu_N = \frac{N\sigma_0^2\bar{x}_N + \sigma^2\mu_0}{N\sigma_0^2 + \sigma^2}$$

and variance

$$\sigma_N^2 = \frac{\sigma^2\sigma_0^2}{N\sigma_0^2 + \sigma^2}$$

where $\bar{x}_N = \frac{1}{N}\sum_{k=1}^{N} x_k$. In the limit of large N, what happens to the mean value μ_N and to the standard deviation σ_N?

4 points

N. Bertschinger
M. Kaschube
V. Ramesh

Machine Learning I
6./7. Exercise Sheet
Due on Friday, Dec 4, 14:15

Exercise 4. *Plot the posterior distribution $p(\mu|X)$ in one graph for various N. The largest N should be at least as large as $N = 500$. Generate data X using a pseudorandom number generator following a Gaussian pdf with mean value equal $\mu = 2$ and variance $\sigma^2 = 4$. The mean value is assumed to be unknown and the prior pdf is also a Gaussian with $\mu_0 = 0$ and $\sigma_0^2 = 8$. Also include the prior in this plot and describe what happens when increasing N.*

2 points

Exercise 5. *The posterior*

$$p(x|X) = \int p(x|\mu)p(\mu|X)d\mu$$

is also Gaussian distributed with mean μ_N and variance $\sigma^2 + \sigma_N^2$. Comment on the result and discuss the large N limit.

2 points

Exercise 6. *Show that the posterior pdf estimate in the Bayesian inference task, for independent variables, can be computed recursively, that is,*

$$p(\theta|x_1, \dots, x_N) = \frac{p(x_N|\theta)p(\theta|x_1, \dots, x_{N-1})}{p(x_N|x_1, \dots, x_{N-1})} .$$

You may either do this for the general expression above or for the example of Gaussian distributions.

2 points

Exercise 7. *Consider the multivariate Gaussian distribution given by*

$$\mathcal{N}(\mathbf{x}|\mu, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^\top \mathbf{\Sigma^{-1}}(\mathbf{x}-\mu)\right\}$$

where μ is a D-dimensional mean vector, $\mathbf{\Sigma}$ is a $D \times D$ covariance matrix, and $|\mathbf{\Sigma}|$ denotes the determinant of $\mathbf{\Sigma}$. By writing the precision matrix (inverse covariance matrix) Σ^{-1} as the sum of a symmetric and an anti-symmetric matrix, show that the anti-symmetric term does not appear in the exponent of the Gaussian, and hence that the precision matrix may be taken to be symmetric without loss of generality. Because the inverse of a symmetric matrix is also symmetric (see Exercise 8), it follows that the covariance matrix may also be chosen to be symmetric without loss of generality.

2 points

N. Bertschinger	**Machine Learning I**
M. Kaschube	6./7. Exercise Sheet
V. Ramesh	Due on Friday, Dec 4, 14:15

Exercise 8. *Show that the inverse of a symmetric matrix is itself symmetric.*

2 points

Bonus exercise. *Show that a real, symmetric matrix of size* $D \times D$ *has* $D(D+1)/2$ *independent parameters.*

2 extra points