

Assignment 1 - Report

Pietro Alovise

11-17-2018

Question 1

The gaussian function is a unimodal distribution, which means that has only one mode and for this particular distribution it coincides with the mean. So in this case we are assuming that value of the deterministic function f for a given \mathbf{x} is the mean value of the distribution of the target. This can be rephrased as assuming a deterministic model $f(\mathbf{x})$ that generates realizations with a random error ε that distributes as $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. Putting everything together we get:

$$\mathbf{t} = f(\mathbf{x}) + \varepsilon$$

A prior observation about the covariance is that we are assuming homoscedasticity, that is the variance of \mathbf{t} is not dependent on the input vector \mathbf{x} .

The spherical covariance matrix means implies two facts:

- All the scalar random variables t_j of the vector \mathbf{t}_i have the same variance σ^2 .
- The fact that the covariance matrix is diagonal means that all the output scalar component t_j of the vector \mathbf{t}_i are independent one another.

Question 2

If we do not assume independence of the samples, we must turn to the joint probability distribution

$$p(\mathbf{T}|f, \mathbf{X}) = p(\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_N|f, \mathbf{X})$$

Question 3

Equation 5 is a linear transformation of a normal distribution which, from its properties, is again a normal distribution equal to:

$$p(\mathbf{t}_i) \sim \mathcal{N}(\mathbf{W} \mathbf{x}_i, \sigma^2 \mathbf{I})$$

Still assuming conditionally independent samples, from 3 the likelihood is just:

$$p(\mathbf{T}|\mathbf{X}, \mathbf{W}) = \prod_{i=1}^N \mathcal{N}(\mathbf{t}_i|\mathbf{W} \mathbf{x}_i, \sigma^2 \mathbf{I})$$

Which we can also write by vectorising the whole, by noting that since all the \mathbf{t}_i have the same variance, the exponents in the probability density function sum up.

$$\begin{aligned} p(\mathbf{T}|\mathbf{X}, \mathbf{W}) &= \mathcal{N}(\mathbf{XW}^T, \mathbf{I}, \sigma^2 \mathbf{I}) = \\ &= \frac{1}{\sigma^2 (2\pi)^{\frac{D}{2}}} \cdot e^{-\frac{1}{2\sigma^2} \sum_i^N (\mathbf{W} \mathbf{x}_i - \mathbf{t}_i)^T (\mathbf{W} \mathbf{x}_i - \mathbf{t}_i)} = \\ &= \frac{1}{\sigma^2 (2\pi)^{\frac{D}{2}}} \cdot e^{-\frac{1}{2\sigma^2} \text{Tr}((\mathbf{XW}^T - \mathbf{T})(\mathbf{XW}^T - \mathbf{T})^T)} \end{aligned}$$

Where we substituted the expression at the exponent $\sum_i^N (\mathbf{W} \mathbf{x}_i - \mathbf{t}_i)^T (\mathbf{W} \mathbf{x}_i - \mathbf{t}_i)$ with $\text{Tr}((\mathbf{XW}^T - \mathbf{T})(\mathbf{XW}^T - \mathbf{T})^T)$ by noting that the summation is just the sum of the diagonal of the matrix $(\mathbf{XW}^T - \mathbf{T})(\mathbf{XW}^T - \mathbf{T})^T$.

Question 4

Using L_1 norm will perform some kind of dimensionality reduction by setting some variables to 0.

The two penalization terms are:

$$p(W) = \frac{1}{\sigma^2 (2\pi)^{\frac{D}{2}}} \cdot e^{-\frac{\text{tr}((W - W_0)(W - W_0)^T)}{2\sigma^2}}$$

$w^T w$: for the L_2 norm $|w|$

Question 5

Question 6

Question 7

Question 8

Question 9

Question 10

Question 11

Question 12

Question 13

Question 14

Question 15

Question 16

Question 17

Question 18

Question 19

Question 20

Question 21

Question 22

Question 23

Question 24