SUSTech

Machine-learning

Fall 2020

Homework 3

刘禹熙 2020 年 10 月 17 日

3.1. Consider a data set in which each data point t_n is associated with a weighting factor $r_n > 9$, so that the sum-of-squares error function becomes

$$E_D(w) = \frac{1}{2}\sum_{n=1}^{N} r_n\{t_n - w^T\phi(x_n)\}^2$$

Find an expression for the solution w^* that minimizes this error function. Give two alternative interpretations of the weighted sum-of-squares error function in terms of (i) data dependent noise variance and (ii) replicated data points.

Solution. 内容...

3.2. We saw in Section 2.3.6 that the conjugate prior for a Gauussian distribution with unknown mean and unknown precision (inverse variance) is a normal-gamma distribution. This property also holds for the case of the conditional Gaussian distribution $p(t|x, w, \beta)$ of the linear regression model. If we consider the likelihood function,

$$p(t|X, w, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|w^T\phi(X_n), \beta^{-1})$$

then the conjugate prior for w and β is given by

$$p(w, \beta) = \mathcal{N}(w|m_0, \beta^{-1}S_0)Gam(\beta|a_0, b_0)$$

Show that the correspondint posterior distribution takes the same functional form, so that

$$p(w, \beta|t) = \mathcal{N}(w|m_N, \beta^{-1}S_N)Gam(\beta|a_N, b_N)$$

and find expressions for the posterior parameters m_N, S_N, a_N, and b_N.

3.3. Show that the integration over w in the Bayesian linear regression model gives the result

$$\int \exp\{-E(w)\}dw = \exp\{E(m_N)\}(2\pi)^{M/2}|A|^{-1/2}$$

Hence show that the log marginal likelihood is given by

$$\ln p(t|\alpha, \beta) = \frac{M}{2} + \frac{N}{2}\ln\beta - E(m_N) - \frac{1}{2}\ln|A| - \frac{N}{2}(2\pi)$$

3.4. Consider real-valued variables X and Y. The Y variable is generated, conditional on X, from the following process:

$$\varepsilon \sim \mathrm{N}(0, \sigma^2)$$
$$Y = aX + \varepsilon$$

where every ε is an independent variable, called a noise term, which is drawn from a Gaussian distribution with mean 0, and standard deviation σ. This is a one-feature linear regression model, where a is the only weight parameter. The conditional probability of Y has distribution $p(Y|X, a) \sim \mathrm{N}(aX, \sigma^2)$, so it can be written as

$$p(Y|X, a) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{1}{2\sigma^2}(Y - aX)^2)$$

Assume we have a training dataset of n pairs (X_i, Y_i) for $i = 1...n$, and σ is known. Derive the maximum likelihood estimate of the parameter a in terms of the training example $X_i^{'}s$ and $Y_i^{'}s$. We recommend you start with the simplest form of the problem:

$$\mathrm{F}(a) = \frac{1}{2}\sum_i (Y_i - aX_i)^2$$

3.5. If a data point y follows the Posson distribution with rate parameter θ, then the probability of a single observation y is

$$p(y|\theta) = \frac{\theta^y e^{-\theta}}{y!}, for\ y = 0, 1, 2, ...$$

You are given data points $y_1, ..., y_n$ independently drawn from a Poisson distribution with parameter θ. Write down the log-likelihood of the data as a function of θ.

3.6. Suppose you are given n obserbations, $X_1, ...X_n$, independent and identically distributed with a Gamma(α, λ) distribution. The following informaztion might be useful for the problem.

(a) If $X \sim \text{Gamma}(\alpha, \lambda)$, then $\mathbb{E}$ [X]=$\frac{\alpha}{\lambda}$ and $\mathbb{E}$ $[X^2]$=$\frac{\alpha(\alpha+1)}{\lambda^2}$

(b) The probabiliuty density function of $X \sim \text{Gamma}(\alpha,\lambda)$ is $f_X(x) = \frac{1}{\Gamma(\alpha)}\lambda^\alpha x^{\alpha-1}e^{-\lambda x}$ where the function Γ is only dependent on α and not λ.

Suppose, we are given a known, fixed value for α. Compute the maximum likelihood estimator for λ.