Homework Ⅲ

Course: Machine Learning(CS405) - Professor: Qi Hao

Question 1

Consider a data set in which each data point t_n is associated with a weighting factor $r_n > 0$, so that the sum-of-squares error function becomes

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

Find an expression for the solution $\mathbf{w}^*$ that minimizes this error function.

Give two alternative interpretations of the weighted sum-of-squares error function in terms of (i) data independent noise variance and (ii) replicated data points.

Question 2

We saw in Section 2.3.6 that the conjugate prior for a Gaussian 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 $p(t|\mathbf{x},\mathbf{w},\beta)$ of the linear regression model. If we consider the likelihood function,

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

then the conjugate prior for $\mathbf{w}$ and β is given by

$$p(\mathbf{w},\beta) = \mathcal{N}(\mathbf{w}|\mathbf{m}_0,\beta^{-1}\mathbf{S}_0)\mathrm{Gam}(\beta|a_0,b_0).$$

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

$$p(\mathbf{w},\beta|\mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N,\beta^{-1}\mathbf{S}_N)\mathrm{Gam}(\beta|a_N,b_N).$$

and find expressions for the posterior parameters $\mathbf{m}_N$, $\mathbf{S}_N$, a_N, and b_N.

Question 3

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

$$\int \exp\{-E(\mathbf{w})\}\mathrm{d}\mathbf{w} = \exp\{-E(\mathbf{m}_N)\}(2\pi)^{M/2}|\mathbf{A}|^{-1/2}.$$

Hence show that the log marginal likelihood is given by

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

Question 4

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

$$\epsilon \sim N(0,\sigma^2)$$

$$Y = aX + \epsilon$$

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 $p(Y|X,a) \sim N(aX,\sigma^2)$, 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:

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

Question 5

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

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

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

Question 6

Suppose you are given n observations, $X_1,\ldots,X_n$, independent and identically distributed with a $Gamma(\alpha,\lambda)$ distribution. The following information might be useful for the problem.

- If $X \sim Gamma(\alpha,\lambda)$, then $\mathbb{E}[X] = \frac{\alpha}{\lambda}$ and $\mathbb{E}[X^2] = \frac{\alpha(\alpha+1)}{\lambda^2}$
- The probability density function of $X \sim Gamma(\alpha,\lambda)$ 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.