ELE 535: Machine Learning and Pattern Recognition
Homework 6

Zachary Hervieux-Moore

Wednesday 14^{th} November, 2018

Exercise 1: The density $f_X(x)$ takes the exponential form

$$f_X(x) = \frac{h(x)}{Z(\theta_0)} e^{\langle \theta_0, t(x) \rangle}$$

You want to use data from m i.i.d. draws from f_X to estimate the value of θ_0. Given the training data, the likelihood function for any parameter value θ is

$$L(\theta) = \prod_{i=1}^{m} \frac{1}{Z(\theta)} h(x_i) e^{\langle \theta, t(x_i) \rangle}$$

The maximum likelihood estimate $\hat{\theta}_0$ of θ_0 is obtained by solving

$$\hat{\theta}_0 = \arg\max_{\theta} L(\theta)$$

Show that $\hat{\theta}_0$ must satisfy

$$\nabla \ln(Z(\hat{\theta}_0)) = \frac{1}{m} \sum_{i=1}^{m} t(x_i)$$

Answer:

Exercise 2: Given m i.i.d. draws from a multivariate Gaussian density, use the method in (Q1) to find the maximum likelihood estimates of the mean μ and covariance matrix Σ of the density.

Answer:

Exercise 3: Empirical statistics, MSE affine prediction, and least squares. Fix a training dataset $\{(x_i, y_i\}_{i=1}^m$, with examples $x_i \in \mathbb{R}^n$ and targets $y_i \in \mathbb{R}^q$. Let X denote the matrix with the examples as its columns, Y denote the matrix with the corresponding targets as its columns. Define the following first and second order empirical statistics of the data:

$$\hat{\mu}_X = \frac{1}{m} X \mathbf{1}_m \quad \hat{\mu}_Y = \frac{1}{m} Y \mathbf{1}_m \tag{1}$$

$$\hat{\Sigma}_X = \frac{1}{m}(X - \hat{\mu}_X \mathbf{1}_m^T)(X - \hat{\mu}_X \mathbf{1}_m^T)^T \quad \hat{\Sigma}_{XY} = \frac{1}{m}(X - \hat{\mu}_X \mathbf{1}_m^T)(Y - \hat{\mu}_Y \mathbf{1}_m^T)^T$$

An optimal MSE affine estimator $\hat{y}(x = W^T x + b$ based on the empirical statistics in (1) must satisfy

$$\hat{\Sigma}_X W = \hat{\Sigma}_{XY} \quad b = \hat{\mu}_Y - W^T \hat{\mu}_X \tag{2}$$

a) Consider the least squares problem

$$W_*, b_* = \underset{W \in \mathbb{R}^{n \times q}, b \in \mathbb{R}^q}{\arg\min} \|Y - W^T X - b \mathbf{1}_m^T\|_F^2 \tag{3}$$

Show that W_*, b_* satisfies (2). Thus directly solving the least squares problem (3) yields an optimal MSE affine estimator for the empirical first and second order statistics in (1).

b) Consider the ridge regression problem

$$W_{r*}, b_{r*} = \underset{W \in \mathbb{R}^{n \times q}, b \in \mathbb{R}^q}{\arg\min} \frac{1}{m} \|Y - W^T X - b \mathbf{1}_m^T\|_F^2 + \lambda \|W\|_F^2, \quad \lambda > 0$$

Determine if W_{r*}, b_{r*} satisfy (2). If not, what needs to be changd in (1) to ensure W_{r*}, b_{r*} satisfy (2). Interpret the change you suggest.

Answer:

Exercise 4: Bias, error covariance, and MSE. Consider random vectors X and Y with a joint density f_{XY} and PD covariance Σ. Let X have mean $\mu_X \in \mathbb{R}^n$ and covariance $\Sigma_X \in \mathbb{R}^{n\times n}$, Y have $\mu_Y \in \mathbb{R}^q$ and covariance $\Sigma_Y \in \mathbb{R}^{q\times q}$, and let the cross-covariance of X and Y be $\Sigma_{XY} \in \mathbb{R}^{n\times q}$.

Let $\hat{y}(x)$ be an estimator of Y given $X = x$, and denote the corresponding prediction error by $E \triangleq Y - \hat{y}(X)$. Of interest is μ_E, Σ_E and the MSE. The estimator is said to be *unbiased* if $\mu_E = \mathbf{0}$.

a) For any estimator $\hat{y}$ with finite μ_E and MSE, show that $\mathrm{MSE}(\hat{y}) = \mathrm{trace}(\Sigma_E) + \|\mu_E\|_2^2$. This shows that the MSE is the sum of two termsLvthe total variance $\mathrm{trace}(\Sigma_E)$ of the error, and the squared norm of the bias $\|\mu_E\|_2^2$.

b) Let $\hat{y}(x) = \mu_Y$. Show that this is an unbiased estimator, determine Σ_E, show that Σ_E is PD, and determine the estimator MSE.

c) The MMSE affine estimator of Y given $X = x$ is

$$\hat{y}^*(x) = \mu_Y + W^{*T}(x - \mu_X) \quad \text{with } \Sigma_X W^* = \Sigma_{XY}$$

Show that $\hat{y}^*(\cdot)$ is an unbiased estimator, determine Σ_E, show that Σ_E is PD, and determine the estimator MSE.

Answer:

Exercise 5: The derivative and gradient of $\|M\|_2$. For $M \in \mathbb{R}^{m\times n}$ define $f(M) = \|M\|_2$. Determine a sufficient condition for the derivative of f to exist at M, and under these conditions find $Df(M)(H)$ and $\nabla f(M)$.

Answer: