ELE 535 - Machine Learning and Pattern Recognition

Fall 2018

HOMEWORK 6: Theory

Q1 The density $f_{\mathrm{X}}(x)$ takes the exponential form

$$f_{\mathrm{X}}(x) = \frac{h(x)}{Z(\theta_0)} e^{<\theta_0, t(x)>}.$$

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

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

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

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

Q2 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.

Q3 **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:

$$\begin{aligned} &\hat{\mu}_{\mathrm{X}} = {}^1\!/_m X \mathbf{1}_m \qquad \hat{\mu}_{\mathrm{Y}} = {}^1\!/_m Y \mathbf{1}_m \\ &\hat{\Sigma}_{\mathrm{X}} = {}^1\!/_m (X - \hat{\mu}_{\mathrm{X}} \mathbf{1}_m^T)(X - \hat{\mu}_{\mathrm{X}} \mathbf{1}_m^T)^T \qquad \hat{\Sigma}_{\mathrm{XY}} = {}^1\!/_m (X - \hat{\mu}_{\mathrm{X}} \mathbf{1}_m^T)(Y - \hat{\mu}_{\mathrm{Y}} \mathbf{1}_m^T)^T \end{aligned} \tag{1}$$

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

$$\hat{\Sigma}_{\mathrm{X}} W = \hat{\Sigma}_{\mathrm{XY}} \qquad b = \hat{\mu}_{\mathrm{Y}} - W^T \hat{\mu}_{\mathrm{X}}. \tag{2}$$

(a) Consider the least squares problem

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

Show that W_*, b_* satisfy (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

$$W_{\mathrm{r}*}, b_{\mathrm{r}*} = \arg \min_{W \in \mathbb{R}^{n \times q}, b \in \mathbb{R}^q} {}^1\!/_m \|Y - W^T X - b \mathbf{1}_m^T\|_F^2 + \lambda \|W\|_F^2, \quad \lambda > 0. \tag{4}$$

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

Q4 **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 $\mu_{\mathrm{X}} \in \mathbb{R}^n$ and covariance $\Sigma_{\mathrm{X}} \in \mathbb{R}^{n \times n}$, Y have mean $\mu_{\mathrm{Y}} \in \mathbb{R}^q$ and covariance $\Sigma_{\mathrm{Y}} \in \mathbb{R}^{q \times q}$, and let the cross-covariance of X and Y be $\Sigma_{\mathrm{XY}} \in \mathbb{R}^{n \times q}$.

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

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

(b) Let $\hat{y}(x) = \mu_{\mathrm{Y}}$. 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 $\mathrm{X} = x$ is

$$\hat{y}^\star(x) = \mu_{\mathrm{Y}} + W^{\star T}(x - \mu_{\mathrm{X}}) \qquad \text{with} \quad \Sigma_{\mathrm{X}} W^\star = \Sigma_{\mathrm{XY}}.$$

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

Q5 **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 $\nabla f(M)$.