Exercise Assignment I (2015.2 - T01)

Deliver Deadline: September, 28th, 2015

1 Polynomial Curve Fitting

Exercise 1) Consider the sum-of-squares error function given by

$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \mathbf{w}) - t_n\}^2$$

in which the function is given by the polynomial

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M.$$

Show that the coefficients $\mathbf{w} = \{w_i\}$ that minimize this error function are given by the solution to the following set of linear equations

$$\sum_{j=0}^{M} A_{ij} w_j = T_i,$$

$$A_{ij} = \sum_{n=1}^{N} (x_n)^{(i+j)} \text{ and}$$

$$T_i = \sum_{n=1}^{N} (x_n)^i t_n.$$

Here a suffix i and j denotes the index of a component, whereas $(x)^i$ x to the power of i.

Exercise 2) Write down the set of coupled linear equations, analogous to

$$\sum_{j=0}^{M} A_{ij} w_j = T_i$$

from the previous exercise, satisfied by the coefficients w_i, which minimize the regularized sum-of-squares error function given by

$$\tilde{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w}) - t_n\}^2 + \frac{\lambda}{2}||\mathbf{w}||^2$$

2 Probability Theory

Exercise 3) Show that if two random variables x and y are independent, then their covariance is zero.

Exercise 4) Using the definition of variance of $f(x)$:

$$\text{var}[f] = \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right].$$

Show that $\text{var}[f]$ satisfies

$$\text{var}[f] = \mathbb{E}\left[f(x)^2\right] - \mathbb{E}[f(x)]^2$$

Exercise 5) Show that the mode (i.e. the maximum) of the Gaussian distribution

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

is given by μ (univariate case). Similarly, show that the mode of multivariate Gaussian is given by $\boldsymbol{\mu}$.

Exercise 6) By setting the derivatives of the log likelihood function

$$\ln\left(\mathbf{x}|\mu,\sigma^2\right) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi)$$

with respect to μ and σ^2 equal to zero, verify the following results:

$$\mu_{\text{ML}} = \frac{1}{N}\sum_{n=1}^{N} x_n,$$

and

$$\sigma^2_{\text{ML}} = \frac{1}{N}\sum_{n=1}^{N} \left(x_n - \mu_{\text{ML}}\right)^2.$$

Where μ_{ML} and σ^2_{ML} are the sample mean and the sample variance respectively.

3 Decision Theory

Exercise 13) Given a loss matrix with elements L_{kj}, the expected risk is minimized if, for each x, we choose the class that minimizes

$$\sum_k L_{kj} p\left(\mathcal{C}_k|\mathbf{x}\right).$$

Verify that, when the loss matrix is given by $L_{kj} = 1 - I_{kj}$, where I_{kj} are the elements of the identity matrix, this reduces to the criterion of choosing the class having the largest posterior probability. What is the interpretation of this form of loss matrix?

Exercise 14) Derive the criterion for minimizing the expected loss when there is a general loss matrix and general prior probabilities for the classes.

Exercise 15) Consider the generalization of the squared loss function

$$\mathbb{E}[L] = \int\int \{y(\mathbf{x}) - t\}^2 p(\mathbf{x}, t) d\mathbf{x} dt$$

for a single target t to the case of multiple target variables described by the vector t, given by

$$\mathbb{E}[L(\mathbf{t}, \mathbf{y}(\mathbf{x}))] = \int\int \|\mathbf{y}(\mathbf{x})\text{-}\mathbf{t}\|^2 p(\mathbf{x}, \mathbf{t}) d\mathbf{x} d\mathbf{t}.$$

Using the claculos of variations, show that the function $\mathbf{y(x)}$ for which this expected loss is minimized is given by $\mathbf{y}(\mathbf{x}) = \mathbb{E}_{\mathbf{t}}[\mathbf{t}|\mathbf{x}]$. Show that this result reduces to

$$y(\mathbf{x}) = \frac{\int tp(\mathbf{x}, t)dt}{p(\mathbf{x})} = \int tp(t|\mathbf{x})dt = \mathbb{E}[t|\mathbf{x}]$$

for the case of a single target variable t.

Exercise 16) By expansion of the square in

$$\mathbb{E}[L(\mathbf{t}, \mathbf{y(x)})] = \int\int \|\mathbf{y(x)\text{-}t}\|^2 p(\mathbf{x}, \mathbf{t})d\mathbf{x}d\mathbf{t},$$

derive a result analogous to

$$\mathbb{E}[L] = \int \{y(\mathbf{x}) - \mathbb{E}[t|\mathbf{x}]\}^2 p(\mathbf{x})d\mathbf{x} + \int \{\mathbb{E}[t|\mathbf{x}] - t\}^2 p(\mathbf{x})d\mathbf{x}.$$

and hence show that the function $\mathbf{y(x)}$ that the minimizes the expected squared loss for the case of a vector $\mathbf{t}$ of target variables is again given by the condicional expectation of $\mathbf{t}$.

4 Information Theory

Exercise 17) Consider an M-state discrete random variable x, and use Jensen's inequality in the form

$$f\left(\sum_{i=1}^{M} \lambda_i x_i\right) \leq \sum_{i=1}^{M} \lambda_i f(x_i))$$

to show that the entropy of its distribution $p(x)$ satisfies $H[x] \leq \ln(M)$.

Exercise 18) Using the definition

$$\mathbf{H}[\mathbf{y}|\mathbf{x}] = -\int\int p(\mathbf{y}, \mathbf{x}) ln\ p(\mathbf{y}|\mathbf{x}) d\mathbf{y} d\mathbf{x}$$

together with the product rule of probability, prove the result

$$\mathbf{H}[\mathbf{x}|\mathbf{y}] = \mathbf{H}[\mathbf{y}|\mathbf{x}] + \mathbf{H}[\mathbf{x}]$$

Exercise 19) Use proof by induction, show that the inequality

$$f\left(\lambda a + (1-\lambda)b\right) \le \lambda f(a) + (1-\lambda) f(b)$$

for convex functions implies the result

$$f\left(\sum_{i=1}^{M} \lambda_i x_i\right) \le \sum_{i=1}^{M} \lambda_i f(x_i)).$$

Exercise 20) Consider two binary x and y having the joint distribution given in

		y	
		0	1
x	0	1/3	1/3
	1	0	1/3

Evaluate the following quantities
(**a**)**H**[**x**]
(**b**)**H**[**y**]
(**c**)**H**[**y**|**x**]
(**d**)**H**[**x**|**y**]
(**e**)**H**[**x**,**y**]
(**f**)**I**[**x**,**y**]
Draw a diagram to show the relationship betweenthese various quantities.

Exercise 21) By applying Jensen's inequality

$$f\left(\sum_{i=1}^{M} \lambda_i x_i\right) \le \sum_{i=1}^{M} \lambda_i f\left(x_i\right))$$

with $f(x) = ln\ x$, show that the arithmetic mean of a set of real numbers is never less than their geometrical mean.