UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Department of Electrical and Computer Engineering

ECE 544NA Pattern Recognition

Homework 1

Fall 2013

Assigned: Thursday, August 29, 2013 Due: Thursday, September 5, 2013

Reading: Bishop Chapter 1

Problem 1.1

Suppose you are given N pairs of points, (x_n, y_y). Your goal is to estimate an $(M-1)^{\text{st}}$-order polynomial function,

$$f(x_n) = \sum_{m=0}^{M-1} a_m x_n^m$$

in order to minimize

$$\epsilon = \sum_{n=0}^{N-1} \|f(x_n) - y_n\|_2^2.$$

The solution to this problem may be expressed in terms of the column vectors $\vec{a} = [a_0, \ldots, a_{M-1}]^T$, $\vec{f} = [f(x_0), \ldots, f(x_{N-1})]^T$, $\vec{y} = [y_0, \ldots, y_{N-1}]^T$, and the matrix

$$X = \begin{bmatrix} 1 & x_0 & \ldots & x_0^{M-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{N-1} & \ldots & x_{N-1}^{M-1} \end{bmatrix}$$

Notice that in this notation, $\epsilon = \|\vec{y} - X\vec{a}\|_2^2$. Suppose that $N \geq M$; prove that ϵ is minimized by $\vec{a}^* = X^\dagger \vec{y}$, where $X^\dagger$ is the pseudo-inverse of X:

$$X^\dagger = (X^T X)^{-1} X^T$$

Problem 1.2

In this problem, your goal is to create an automatic banana detector. You have a bunch of fruit coming down the conveyor belt; each piece of fruit passes in front of an old-school electric eye, which takes just one measurement of the brightness of the fruit, x. Under these barbaric circumstances, the best fruit classifier you can create is one that compares x to a threshold θ. Whenever $x \geq \theta$, you conclude that the fruit is a banana; whenever $x < \theta$, you conclude that the fruit is an apple. Let ω be the true class of the fruit, and define the prior probabilities $\pi_A = \Pr\{\omega = A\}$, and $\pi_B = 1 - \pi_A = \Pr\{\omega = B\}$. Let $\hat{\omega}$ be the estimated class of the fruit; your decision rule is

$$\hat{\omega} = \begin{cases} A & x < \theta \\ B & x \geq \theta \end{cases}$$

On a particular brightness scale, apples (class A) have a brightness distributed as

$$p(x|\omega = A) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \tag{1}$$

and bananas (class B) have a brightness distributed as

$$p(x|\omega = B) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-d')^2} \tag{2}$$

where d' is defined to be the normalized difference between the means of the two distributions.

In the following derivations, you will certainly find the Φ-function and Q-function useful; they are:

$$\Phi(x) = 1 - \Phi(-x) = Q(-x) = 1 - Q(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\xi^2} d\xi$$

(a) Since this is a banana-detector, any false banana detections are called "false alarms." The false alarm probability is therefore defined as

$$P_{FA}(d', \theta) = \Pr\{\hat{\omega} = B | \omega = A\}$$

Express P_{FA} as a function of d' and θ.

(b) The missed-detection probability is

$$P_{MD}(d', \theta) = \Pr\{\hat{\omega} = A | \omega = B\}$$

Express P_{MD} as a function of d' and θ.

(c) There are many different ways to summarize the performance of a signal detection system. One useful metric is the equal error rate, defined as

$$P_{EER}(d') = P_{FA}(d', \theta) = P_{MD}(d', \theta)$$

at the particular value of θ chosen so that $P_{MD} = P_{FA}$. Find P_{EER} of the banana-detector, as a function of d'. Sketch the function $P_{EER}(d')$ by hand, showing the values achieved at $d' = 0$, $d' = 1$, and $d' \to \infty$.

(d) For any particular setting of the system, the total error rate is defined to be

$$P_E(d', \theta, \pi_A, \pi_B) = \Pr\{\hat{\omega} \neq \omega\}$$

Find P_E as a function of d', θ, π_A and π_B.

(e) The Bayes risk is defined to be the lowest error rate that can be achieved for a particular set of priors and likelihoods, and θ^* is the value of the system parameter θ that achieves the Bayes risk, thus

$$P_{BR}\left(d', \frac{\pi_A}{\pi_B}\right) = \min_{\theta} \Pr\{\hat{\omega} \neq \omega\}$$

$$\theta^* = \arg\min_{\theta} \Pr\{\hat{\omega} \neq \omega\}$$

Find P_{BR} and θ^* as functions of d' and $\frac{\pi_A}{\pi_B}$. Write a few sentences of intelligible English sufficient to convince me that, in the case $\pi_A = \pi_B$, your formula for θ^* is intuitively obvious.

Matlab Exercises

Problem 1.3

Write a script in matlab that generates eleven values of t, uniformly distributed in $0 \leq t < 1$ (use `rand`), then generates eleven corresponding points from a noisy sine wave:

$$y_t = z_t + v_t$$

where $z_t = \sin(2\pi t)$, and v_t are i.i.d. Gaussian random variables with standard deviation 0.1 ($\sigma^2 = 0.01$; use `randn`). Use the pseudo-inverse method described in problem 1 to find the coefficients of best-fitting polynomials of oder $M-1=1$, $M-1=3$, and $M-1=10$.

Desert Island Alert: If you use the matlab functions `polyfit` or `pinv`, or their equivalents, anywhere in your code, your answer will be marked as wrong. The purpose of this problem is not to teach you how to use matlab functions, it is to teach you how to *replace* matlab functions, in case you're ever stranded on a desert island without access to matlab.

Hand in your code (at most one page), and three figures (preferably all pasted onto the same one page) showing the following functions for a first-order, third-order, and tenth-order polynomial, respectively. Each figure should show the noiseless sine wave z_t (plotted using a dashed line), the eleven samples of the noisy sine wave y_t (plotted using 'x' points), and the recovered polynomial function $f(t)$ (plotted using a solid line). Note: in order to generate the smooth plots of z_t and $f(t)$, you will need to define a fine-grained uniformly-sampled time vector such as `tau=[0:0.01:1]`, then if you have previously defined the functions z() and f(), you can enter something like `plot(tau,z(tau),'--');` or `plot(tau,f(tau,a),'-');`.

Problem 1.4

A detection-error tradeoff (DET) curve is a plot of the points $P_{MD}(\theta)$ versus $P_{FA}(\theta)$ that can be achieved, by varying θ, given a particular set of prior probabilities and likelihoods.

Assume that the brightness likelihoods of apples and bananas are as given in problem 2, thus

$$p(x|A) = \mathcal{N}(0,1) \qquad p(x|B) = \mathcal{N}(d',1)$$

For each of the following settings of d' and N, use the `randn` function to generate N random apples (samples of x_{nA} drawn at random from $p(x_{nA}|A)$), and N random bananas (samples of x_{nB} drawn from $p(x_{nB}|B)$). Define the empirical false-alarm and missed-detection probabilities to be

$$\hat{P}_{FA} = \frac{1}{N}\sum_{n=1}^{N}[x_{nA} \geq \theta]$$
$$\hat{P}_{MD} = \frac{1}{N}\sum_{n=1}^{N}[x_{nB} < \theta]$$

where the $[P_0]$ function is defined by

$$[P_0] = \begin{cases} 1 & \text{Proposition } P_0 \text{ is true} \\ 0 & \text{Proposition } P_0 \text{ is false} \end{cases}$$

For each of the following settings, plot an empirical DET curve: a single plot showing all values of $\hat{P}_{FA}$, $\hat{P}_{MD}$ that can be achieved for any possible value of θ. In each case, label the equal error rate, or the closest thing to an equal error rate that is empirically achieved. Hand in your code (at most one page), and four figures for the four subsections (preferably all pasted onto the same page).

(a) $d'=1$, $N=4$ (b) $d'=1$, $N=1000$
(c) $d'=2$, $N=4$ (d) $d'=2$, $N=1000$