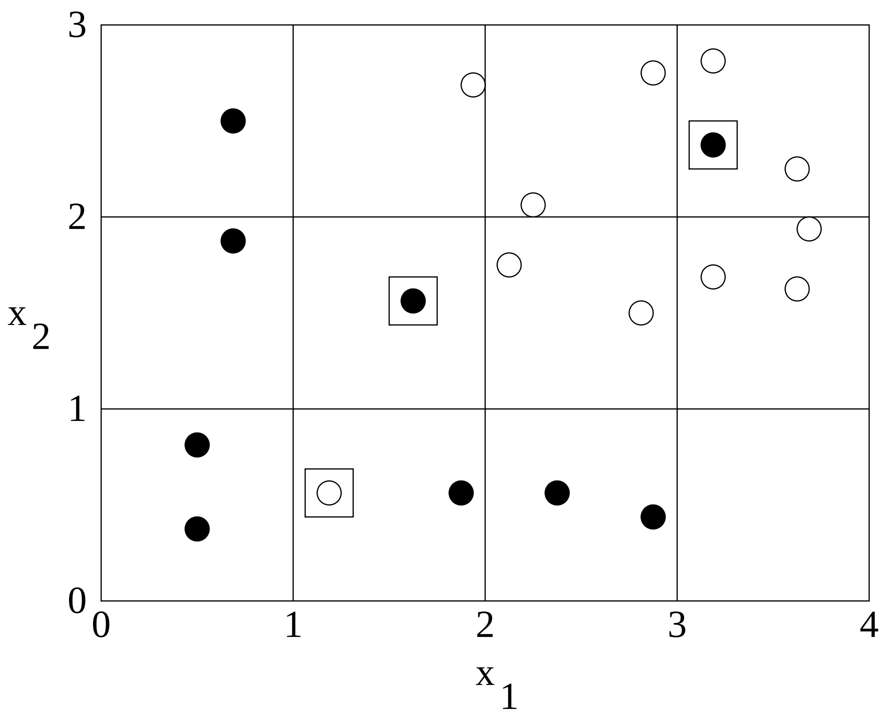
ANSWERS FOR FIRST TEST

1/30	
2/25	
3/25	
4/20	
T/100	

STA 414/2104 — First Test — 2007-02-06

No books, notes, or calculators are allowed.

Question 1: [30 marks] Consider a classification problem in which there are two real-valued inputs, x_1 and x_2, and a binary (0/1) target (class) variable, t. There are 20 training cases, plotted below. Cases where $t = 1$ are plotted as black dots, cases where $t = 0$ as white dots, with the location of the dot giving the inputs, x_1 and x_2, for that training case.



a) Estimate the error rate of the one-nearest-neighbor (1-NN) classifier for this problem using leave-one-out cross validation. (Ie, cross validation in which each training case is predicted using all the other training cases.)

Three training cases will be mis-classified using 1-NN, based on the other training cases. They are marked above. The leave-one-out cross-validation error rate is therefore 3/20.

b) Suppose we use the three-nearest-neightbor (3-NN) method to estimate the probability that a test case is in class 1. For test cases with each of the following sets of input values, find the estimated probability of class 1.

$x_1 = 1,\ x_2 = 1$

Two of the three training cases nearest to this point are in class 1, so the estimated probability of class 1 is 2/3.

$x_1 = 2,\ x_2 = 2$

One of the three training cases nearest to this point are in class 1, so the estimated probability of class 1 is 1/3.

c) Suppose we use 3-NN to estimate class probabilities, and that our loss function is $L_{00} = L_{11} = 0$, $L_{01} = 1$, and $L_{10} = 3$, where L_{kj} is the loss if we classify a case as class j when it is actually class k. Say how we should classify each of the following test points, and give the expected loss when we classify this way.

$x_1 = 1,\ x_2 = 1$

If we classify this point in class 1, the expected loss will be $P(\textit{class } 0)L_{01} = (1/3) \cdot 1 = 1/3$.

If we classify this point in class 0, the expected loss will be $P(\textit{class } 1)L_{10} = (2/3) \cdot 3 = 2$.

We should therefore classify it as class 1, with expected loss $1/3$.

$x_1 = 2,\ x_2 = 2$

If we classify this point in class 1, the expected loss will be $P(\textit{class } 0)L_{01} = (2/3) \cdot 1 = 2/3$.

If we classify this point in class 0, the expected loss will be $P(\textit{class } 1)L_{10} = (1/3) \cdot 3 = 1$.

We should therefore classify it as class 1, with expected loss $2/3$.

Question 2: [25 marks] Let X_1, X_2, X_3, ... for a sequence of binary (0/1) random variables. Given a value for θ, these random variables are independent, and $P(X_i = 1) = \theta$ for all i. Suppose that we are sure that θ is at least 1/2, and that our prior distribution for θ for values 1/2 and above is uniform on the interval $[1/2, 1]$. We have observed that $X_1 = 0$, but don't know the values of any other X_i.

a) Write down the likelihood function for θ, based on the observation $X_1 = 0$.

$L(\theta) \;=\; P(X_1 = 0 \,|\, \theta) \;=\; 1 - \theta$

b) Find an expression for the posterior probability density function of θ given $X_1 = 0$, simplified as much as possible, with the correct normalizing constant included.

The prior density is $P(\theta) = 2$ *for* $\theta \in [1/2, 1]$, *0 otherwise.*

The posterior density is $P(\theta \,|\, X_1 = 0) \;=\; 0$ *for* $\theta \notin [1/2, 1]$, *and otherwise* $P(\theta \,|\, X_1 = 0) \;\propto\; P(\theta)\, L(\theta) \;\propto\; 2\,(1-\theta)$. *The normalizing constant can be found by evaluating* $\int_{1/2}^{1} 2\,(1-\theta)\, d\theta \;=\; 1/4$, *from which we find that* $P(\theta \,|\, X_1 = 0) \;=\; 8\,(1-\theta)$.

c) Find the predictive probability that $X_2 = 1$ given that $X_1 = 0$. An actual number is required.

$P(X_2 = 1 \,|\, X_1 = 0) \;=\; \int P(X_2 = 1 \,|\, \theta)\, P(\theta \,|\, X_1 = 0)\, d\theta \;=\; = \;\int_{1/2}^{1} \theta\, 8\,(1-\theta)\, d\theta \;=\; 2/3$

d) Find the probability that $X_2 = X_3$ given that $X_1 = 0$. An actual number is required.

$$
\begin{aligned}
P(X_2 = X_3 \,|\, X_1 = 0) \;&=\; \int P(X_2 = X_3 \,|\, \theta)\, P(\theta \,|\, X_1 = 0)\, d\theta \\
&=\; \int [\, P(X_2 = 0, X_3 = 0 \,|\, \theta) + P(X_2 = 1, X_3 = 1 \,|\, \theta) \,]\, P(\theta \,|\, X_1 = 0)\, d\theta \\
&=\; \int [\, P(X_2 = 0 \,|\, \theta) P(X_3 = 0 \,|\, \theta) + P(X_2 = 1 \,|\, \theta) P(X_3 = 1 \,|\, \theta) \,]\, P(\theta \,|\, X_1 = 0)\, d\theta \\
&=\; \int_{1/2}^{1} ((1-\theta)^2 + \theta^2)\, 8\,(1-\theta)\, d\theta \\
&=\; 7/12
\end{aligned}
$$

Note that X_2 *and* X_3 *are independent given* θ, *but they are not independent given just* X_1.

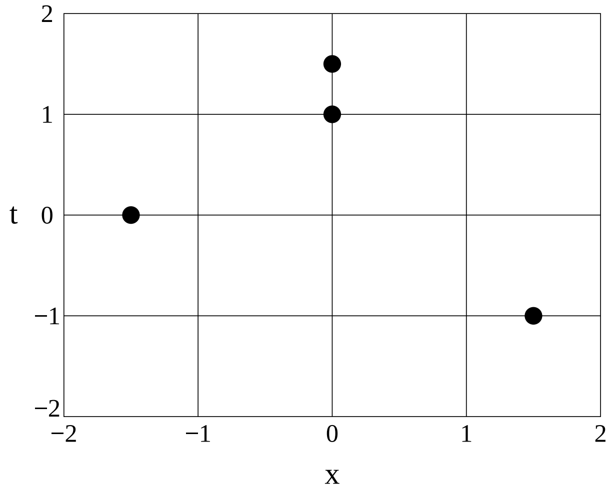
Question 3: [25 marks] Consider a linear basis function regression model, with one input and the following three basis functions:

$$\begin{aligned}
\phi_0(x) &= 1 \\
\phi_1(x) &= x \\
\phi_2(x) &= \begin{cases} 1 - x^2 & \text{if } |x| < 1 \\ 0 & \text{if } |x| \geq 1 \end{cases}
\end{aligned}$$

The model for the target variable, t, is that $P(t|x,w) = N(t|y(x,w),1)$, where

$$y(x,w) = \sum_{j=0}^{M-1} w_j \phi_j(x)$$

Suppose we have four data points, as plotted below:



What is the maximum likelihood (least squares) estimate for the parameters w_0, w_1, and w_2? Elaborate calculations should not be necessary.

Note that $\phi_2(x)$ is zero for the data points where $x = -1.5$ and $x = +1.5$. So the value of w_2 will not affect the value of $y(x,w)$ at these points. It can therefore be used to fit the two data points at $x = 0$ (where $\phi(x) = 1$) as well as possible, regardless of what w_0 and w_1 are. This in turn means that we can use w_0 and w_1 to fit the two data points at $x = -1.5$ and $x = +1.5$. Looking at the line joining these two points, we see that the intercept is $-1/2$ and the slope is $-1/3$. We will therefore fit these points exactly if we use $w_0 = -1/2$ and $w_1 = -1/3$. Choosing $w_2 = 1.75$ will then lead to $y(0,w) = 1.25$, which is the best value we can have for fitting the two data points at $x = 0$.

Question 4: [20 marks] Consider the Poisson model, with unknown positive parameter λ, for a random variable, X, that takes on non-negative integer values:

$$P(X = x) \;=\; \frac{\lambda^x}{x!}\exp(-\lambda)$$

Show how this model can be expressed in the exponential family form, with a natural parameter η, a sufficient statistic $u(x)$, and functions $h(x)$ and $g(\eta)$, so that the probability for a value x has the form

$$P(X = x) \;=\; h(x)g(\eta)\exp(\eta^T u(x))$$

We can let $\eta = \log\lambda$ *and* $u(x) = x$. *The probability function can then be written as*

$$P(X = x) \;=\; (1/x!)\exp(-\exp(\eta))\exp(\eta u(x))$$

so $h(x) = 1/x!$ *and* $g(\eta) = \exp(-\exp(\eta))$.