UNIVERSITY OF EDINBURGH

FACULTY OF SCIENCE AND ENGINEERING

LFD1

Date: ?? April 2001 DRAFT **Time: 09:00-10:30**

DRAFT

This will describe the degree (MSc/AI4,etc.)

Examiners:	Name of external examiner	(External)
	Name of exam board chair	(Chair)

INSTRUCTIONS TO CANDIDATES

Answer TWO questions.

If you attempt three questions, cross out one answer; if you do not, then the examiners will cross out the last one you answered.

Each complete question carries equal weight and is marked out of 100. The parts of a question may not all be worth the same amount; the marks at the side of the questions indicate how these will normally be apportioned.

Write as legibly as possible.

THIS EXAMINATION WILL BE MARKED ANONYMOUSLY

LFD1

1. Consider a set of N-dimensional data $\boldsymbol{x}^\mu, \mu = 1, \ldots, P$. Each datapoint $\boldsymbol{x}^\mu$ has a corresponding class label, c^μ.

   (a) Explain how the K nearest neighbour method (KNN) can be used to classify an unlabelled test data point $\boldsymbol{x}^*$. [20%]

   (b) Describe how to reduce the above training data to M-dimensional data using Principal Components Analysis (PCA). [20%]

   (c) Consider a dataset in two dimensions where the data lies on the circumference of a circle of unit radius. What would be the effect of using PCA on this dataset, in which we attempt to reduce the dimensionality to 1? Suggest an alternative one dimensional representation of the data. [35%]

   (d) Consider two vectors $\boldsymbol{x}^a$ and $\boldsymbol{x}^b$ and their corresponding PCA approximations $\boldsymbol{c} + \sum_{i=1}^M a_i \boldsymbol{e}^i$ and $\boldsymbol{c} + \sum_{i=1}^M b_i \boldsymbol{e}^i$, where the eigenvectors $\boldsymbol{e}^i, i = 1, \ldots M$ are mutually orthogonal and have unit length. The eigenvector $\boldsymbol{e}^i$ has corresponding eigenvalue λ^i.

   Approximate $(\boldsymbol{x}^a - \boldsymbol{x}^b)^2$ by using the PCA representations of the data, and show that this is equal to $(\boldsymbol{a} - \boldsymbol{b})^2$. [25%]

2. Consider data $\{(\boldsymbol{x}^i, c^i), i = 1, \ldots, n_1 + n_2\}$. This data consists of N-dimensional examples $\boldsymbol{x}^i$, each with a corresponding class label, $c^i = 1$ or $c^i = 2$. There are n_1 training examples from class 1, and n_2 training examples from class 2.

   (a) Explain Bayes' rule, and how it can be used to form a classifier based on $p(c)$ and $p(x|c)$. [15%]

   Show that the value of x that represents the decision boundary of the classifier satisfies the expression

$$\log \frac{p(x|c=2)}{p(x|c=1)} = \log \frac{p(c=1)}{p(c=2)}$$

   [25%]

   What is the Naive Bayes classification method, and why is it naive? Discuss a situation in which the assumption used in Naive Bayes is incorrect. [25%]

   (b) The Gaussian distribution in one dimension is defined as

$$p(x|c) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

   We decide to fit a Gaussian to each class from a dataset of one-dimensional data. Show that the Maximum Likelihood estimator of μ_1 is $\hat{\mu}_1 = \frac{1}{n_1} \sum_{x \in class1}^{n_1} x$ and that the ML estimate of σ_1^2 is $\hat{\sigma}_1^2 = \frac{1}{n_1} \sum_{x \in class1}^{n_1} (x - \hat{\mu}_1)^2$ [35%]

3. (a) WowCo.com is a new startup prediction company. After years of failures, they eventually find a neural network with a trillion hidden units that achieves zero test error on every learning problem posted on the internet up till January 2002. Each learning problem included a training and test set. Proud of their achievement, they market their product aggressively with the claim that it 'predicts perfectly on all known problems'. Would you buy this product? Justify your answer. [35%]

(b) Consider iid training data $D = \{(\boldsymbol{x}^\mu, c^\mu), \mu = 1, \ldots, P\}$, $c^\mu \in \{0, 1\}$. Define logistic regression, and comment on the decision boundary. [20%]

Show that, using logistic regression, the log likelihood of the training data is

$$L(\mathbf{w}) = \sum_{\mu} c^\mu \log \sigma\left(\theta + \mathbf{w}^T \mathbf{x}^\mu\right) + (1 - c^\mu) \log\left(1 - \sigma\left(\theta + \mathbf{w}^T \mathbf{x}^\mu\right)\right)$$

where $\sigma(x) = 1/(1 + \exp(-x))$. [15%]

Describe gradient ascent and how it can be used to train the logistic regression model. [15%]

How would you calculate the boundary $p(c = 1|x) = 0.9$? [15%]