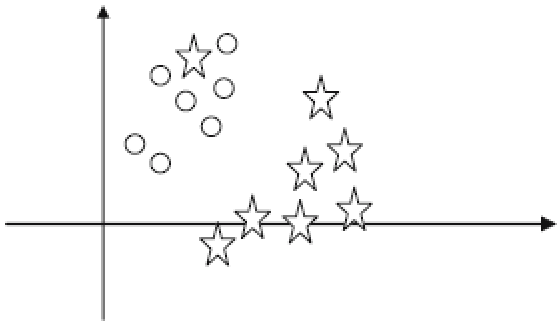
Exam 26 January 2015 **Duration: 2h30min**

1.

a) Consider a binary classifier, where the discriminant function is given by $f(x_1, x_2) = 6.5 + 0.5x_1 - 0.9x_2 + 0.1x_1x_2$. Check is the two observations (1.2; 75) and (10.4;-15.0) are predicted to belong to the same class or not.

b) Assume now that you are given the training set depicted in the figure to train a binary classifier. Is the 1-NN (nearest neighbour with k=1) a good option? What classifier would you suggest?

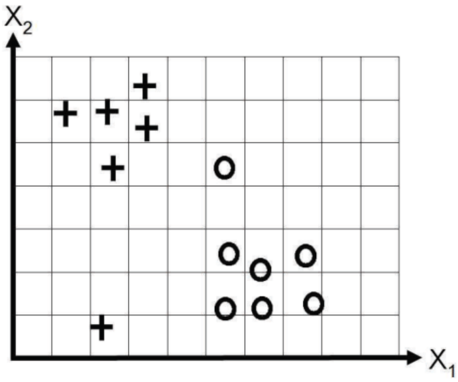


2. In many applications, the classifier is allowed to “reject” a test example rather than classifying it into one of the classes. Consider, for example, a case in which the cost of a misclassification is \$10 but the cost of having a human manually make the decision is only \$3. We can formulate this as the following loss matrix:

decision $\hat{y}$	true label y	
	0	1
predict 0	0	10
predict 1	10	0
reject	3	3

a) Suppose P(y = 1|x) is predicted to be 0.2. Which decision minimizes the expected loss?
b) Now suppose P(y = 1|x)=0.4. Now which decision minimizes the expected loss?
c) Show that in general, for this loss matrix, but for any posterior distribution, there will be two thresholds θ_0 and θ_1 such that the optimal decision is to predict 0 if $p_1 < \theta_0$, reject if $\theta_0 \le p_1 \le \theta_1$, and predict 1 if $p_1 > \theta_1$ (where $p_1 = p(y = 1|x)$). What are these thresholds?

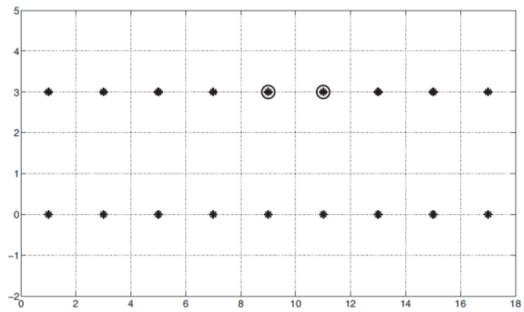
3. Consider the (training) data in the figure, where we fit a linear model $w_0+w_1x_1+w_2x_2$. Suppose we fit the model by maximum likelihood, i.e., we minimize $J(w) = -L(w,D_{train})$, where $L(w,D_{train})$ is the log likelihood on the training set.



EXTRA FIGURES AT THE END OF THE EXAM

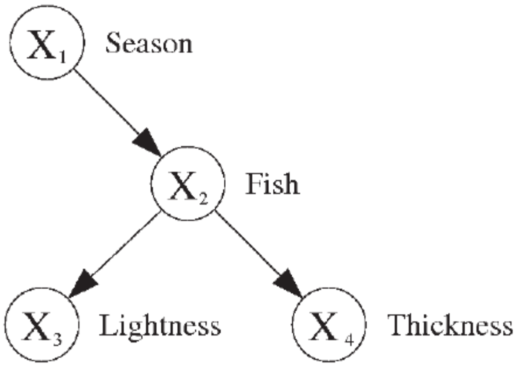
a) Sketch a possible decision boundary corresponding to learnt model. Is your answer (decision boundary) unique? How many classification errors does your method make on the training set?
b) Now suppose we regularize only the w_0 parameter, i.e., we minimize $J_0(w) = -L(w,D_{train}) + \lambda w_0^2$. Suppose λ is a very large number, so we regularize w_0 all the way to 0, but all other parameters are unregularized. Sketch a possible decision boundary. How many classification errors does your method make on the training set?
c) Now suppose we heavily regularize only the w_1 parameter, i.e., we minimize $J_1(w) = -L(w,D_{train}) + \lambda w_1^2$. Sketch a possible decision boundary. How many classification errors does your method make on the training set?
d) Now suppose we heavily regularize only the w_2 parameter. Sketch a possible decision boundary. How many classification errors does your method make on the training set?

4. In the Figure, we show some data points which lie on the integer grid. (Note that the x-axis has been compressed; distances should be measured using the actual grid coordinates.)



Suppose we apply the K-means algorithm to this data, using $K = 2$ and with the centers initialized at the two circled data points. Draw the final clusters obtained after K-means converges (show the approximate location of the new centers and group together all the points assigned to each center).

5. Consider the Bayes net shown in Figure.



Here, the nodes represent the following variables
$X_1 \in$ {winter, spring, summer, autumn},
$X_2 \in$ {salmon, sea bass}
$X_3 \in$ {light, medium, dark},
$X_4 \in$ {wide, thin}
The corresponding conditional probability tables are

$$p(x_2|x_1) = \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \\ 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix} \quad p(x_3|x_2) = \begin{bmatrix} 0.33 & 0.33 & 0.34 \\ 0.8 & 0.1 & 0.1 \end{bmatrix} \quad p(x_4|x_2) = \begin{bmatrix} 0.4 & 0.6 \\ 0.95 & 0.05 \end{bmatrix}$$

a) Suppose the fish was caught on December 20 — the end of autumn and the beginning of winter — and thus let $p(x_1) = [0.5\ \ 0\ \ 0\ \ 0.5]$. Suppose it is known that the fish is thin and light. Classify the fish as salmon or sea bass.

6. Let $x \in \{0, 1\}$ denote the result of a coin toss ($x = 0$ for tails, $x = 1$ for heads). The coin is potentially biased, so that heads occurs with probability θ_1. Suppose that someone else observes the coin flip and reports to you the outcome, y. But this person is unreliable and only reports the result correctly with probability θ_2; i.e., $p(y|x, \theta_2)$ is given by

	y = 0	y = 1
x = 0	θ_2	$1 - \theta_2$
x = 1	$1 - \theta_2$	θ_2

Assume that θ_2 is independent of x and θ_1.

a) Write down the joint probability distribution $p(x, y|\theta)$ as a 2 × 2 table, in terms of $\theta = (\theta_1, \theta_2)$.
b) Suppose have the following dataset: x = (1, 1, 0, 1, 1, 0, 0), y = (1, 0, 0, 0, 1, 0, 1). What are the MLEs for θ_1 and θ_2? Justify your answer. Hint: note that the likelihood function factorizes, $p(x, y|\theta) = p(y|x, \theta_2)p(x|\theta_1)$
c) What is $p(D|\theta,M_2)$ where M_2 denotes this 2-parameter model? (You may leave your answer in fractional form if you wish.)

Name:

EXTRA FIGURES FOR GROUP 3