UN3106 – Sample Exam Problems

The sample questions below were all actual exam questions in previous terms and should give you a fairly accurate idea of the kind of questions we ask.

Solutions: Some of these problems are quite similar to homework problems. I will not provide solutions to the remaining problems, simply because I do not want you to overfit on this sample.

Problem 1: Short questions

Short answers (about one sentence) are sufficient.

(a) (**Yes/No**) Does the Bayes-optimal classifier depend on the choice of loss functions?

(b) Is a m-nearest neighbor method equally prone to overfit for all values of m?

(c) What characterizes supervised (as opposed to unsupervised) learning?

From the homework

The following homework problems are former exam problems:

- Problem 1 on HW 1 (naive Bayes)
- Problem 2 on HW 2 (cross validation)

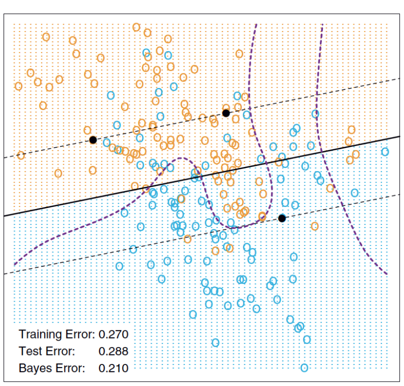
You should review all homework problems except those involving R. The exam will not involve any form of coding problems.

Problem 2: Classification algorithms

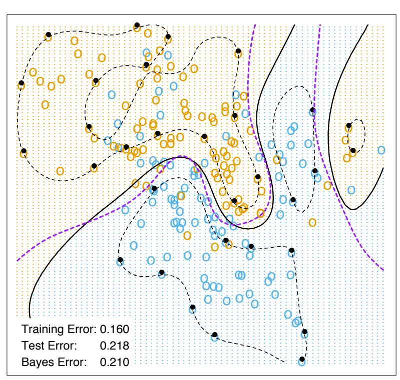
Can you give an example of a data distribution for which a maximum margin classifier would not be optimal? Please define or describe the class-conditional distributions of your example distribution.

Problem 3: Decision boundaries

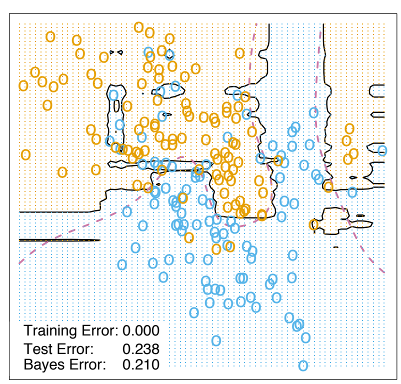
Note: *Figure (ii) and (iii) below refer to classifiers we have not discussed in class, and which would of course not show up in the midterm. I have included the question anyway for illustration.* The following pictures, which we have all seen in class, show the output of several different classifiers. Recall that the thick line is the decision boundary determined by the classifier; you can ignore the dashed lines.



(a)



(b)



(c)

For each of the three pictures:

- Name at least one classifier which could have produced this solution. Explain why.
- Name at least one classifier which could not have produced the solution. Explain why not.

Problem 4: Jumping the queue with improvable machine learning

London's Heathrow Airport operates an "automatic border agent" that checks passports: The system scans the traveler's face, compares it to biometric information stored in the passport, and reports "match" or "no match".

(a) Phrase the passport check as a classification problem: Suppose the face scan is represented as a point $\mathbf{a} \in \mathbb{R}^k$ and the biometric information in the passport as $\mathbf{b} \in \mathbb{R}^m$. Assume that *both* $\mathbf{a}$ *and* $\mathbf{b}$ are new data that the system has not encountered before (that is, the system does not know $\mathbf{b}$ for every passport in the world beforehand). Please specify the sample space on which your classifier is defined, how many classes there are, and the set of class labels.

(b) In its early days, the Heathrow system was popular among European travelers, since it almost always reports "no match". (The traveler is then taken straight to an immigration officer and has effectively saved time by jumping the queue.) What does that tell you about the loss function that was used to train the system?

(c) Considering that most of us usually travel with a legitimate passport, can you think of a method to improve the accurracy of the system, without having to improve the quality of the data? Please name a method and explain why you consider it appropriate.

Problem 5: Maximum margin classifier

Consider a two-class classification problem, with data in $\mathbb{R}^d$. Show that two training data points, one in each class, are sufficient to completely determine a maximum-margin classifier, regardless of the value of d.
Hint: You do not need to give any equations. A picture and an explanation of how the two points determine the affine hyperplane is sufficient as an answer.

Problem 6: Linear classification

Consider a perceptron classifier in $\mathbb{R}^2$, given by a hyperplane with orthogonal vector v_{H} and offset c, and two points $\mathbf{x}_1$ and $\mathbf{x}_2$. Suppose that

$$v_{\text{H}} := \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \qquad c := \frac{1}{2\sqrt{2}} \qquad \mathbf{x}_1 := \begin{pmatrix} -3 \\ 0 \end{pmatrix} \qquad \mathbf{x}_2 := \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} .$$

(a) Compute the classification result for $\mathbf{x}_1$ and $\mathbf{x}_2$.

(b) If the classifier is given by the same pair $(v_{\text{H}}, -c)$, but was trained as an SVM, do the results change? Please explain your answer.

(c) Which cost function does the perceptron cost function approximate, and why do we approximate it?