Intelligent Interaction Pearl 2020-2021
Sample exam questions (with answers)

Evaluation of performance

Assume we have trained a classifier on a training set, to recognize samples from class C1 and C2, and evaluated it on an independent test set. The following confusion matrix summarizes the results. Your task is to calculate different performance measures from the confusion matrix.

		Predicted class	
		C1	**C2**
True (ground truth)	**C1**	300	75
	C2	25	200

1. *What is the Accuracy?*
2. *What is the Precision for class C1?*
3. *What is the Recall for class C1?*
4. *What is the Precision for class C2?*
5. *What is the Recall for class C2?*

Answers

1. Accuracy : 0.83

2. and 3.

For questions 2 and 3, we consider C1 as the positive class (and C2 as negative). The confusion matrix reads as (notice what we consider TP, TN, FP, FN):

		Predicted class	
		C1	**C2**
True (ground truth)	**C1**	300 (TP)	75 (FN)
	C2	25 (FP)	200 (TN)

Precision for C1 = TP/(TP+FP) = 300 / (300 + 25) = 0.92

Recall for C1 = TP/(TP+FN) = 300 / (300 + 75) = 0.8

4. and 5.

For questions 4 and 5, we consider C2 as the positive class (and C1 as negative). The confusion matrix reads as (notice what we consider TP, TN, FP, FN):

		Predicted class	
		C1	**C2**
True (ground truth)	**C1**	300 (TN)	75 (FP)
	C2	25 (FN)	200 (TP)

Precision for C1 = TP/(TP+FP) = 200 / (200 + 75) = 0.73

Recall for C1 = TP/(TP+FN) = 200 / (200 + 25) = 0.89

Recognition

A robot that navigates in a garden is equipped with a camera system to see what is in front of it. The robot is programmed to recognize roses, iris and violets, pick them up and store in different places. The term 'recognition' applies to this process (of recognizing different types of flower) but this term is rather general. Another, more specific term can be used here. From the following list, choose the term which best describes this type of pattern recognition.

a) Classification
b) Verification
c) Clustering
d) Identification
e) Detection
f) Authorization/Authentication

The correct answer is **a) classification.**

Comments:
The robot recognizes a flower and assigns to it one of the three possible classes (roses, iris, violets) – this is a classification task as the robot decides which class the sample belongs to.

Model validation

You have trained a classification model and you observe that on your training data, you achieve a very low error (~1%). When you test your model on an independent test set, you achieve a test error of 25%. What do you infer from these results?

a) The model overfitted the training data
b) The model has good generalization
c) The classifier is good: 75% on new data in a great result
d) The error on the training set is very low: the classifier is robust

The correct answer is **a) The model overfitted the training data.**

Comments:
When your classifier makes a very small error on the training data, it means that it is so complex to learn all the details of the distribution of the training data. If the error on the test data is much bigger than that on the training data, this indicates that the classifier is not able to make reliable predictions on the test data, and it is too specialized to recognize training data. It thus overfits to the training data.

A different case is when the training error is low, and also the error made on the classification of test data is low. In this case, we say that it generalizes well to test data (which is new data, completely independent of the training data).

Discriminant functions

What is a discriminant function?

a) a function that discriminate data samples from a data set
b) a function that models the distribution of the samples in a given class, and computes a score for new samples to belong to that class
c) a function that computes the Prior probability of a class
d) a classifier that predicts the class of new samples
e) a function that models the probability distribution of new data

The correct answer is **b)**

Comments:

As explained in lecture 2, and repeated in lecture 03, a discriminant function is defined to represent the distribution of the data points (samples) of one class.
A classifier uses a discriminant functions $\mathbf{g_i(x)}$ to compute the score (for example, the probability) that a new test sample belongs to the class ***i.***

Discriminant function (computation) and classification

Let us have two classes, that contains data samples that can be modeled by a one-dimensional Normal distribution.

We computed the mean and variance of the samples in the two classes, and their values are:

Class 1: mean m_1 = -4, variance s_1^2= 10

Class 2: mean m_2 = 5, variance s_2^2 = 10

The two distributions have same prior probability P(Class 1) = P(Class 2) = 0.5.

a) Compute the value of the decision criterion x*.

b) In which class do you classify the sample x_1 = 0.9?

c) In which class do you classify the sample x_2 = 0 ?

d) In which class do you classify the sample x_3 = 3 ?

e) In which class do you classify the sample x_4 = -2 ?

a) *decision criterion is x* = 0.5*

solution:

We can represent the two classes using two discriminant functions $g_1(x)$ and $g_2(x)$, and choose them to be the posterior probability of classifying a sample in one of the two classes. Thus, $g_1(x)$ = P(Class1 | x) and $g_2(x)$ = P(Class2 | x).

As the classes are normally distributed, we can represent their class-conditional probability as a Gaussian functions with the given mean and variance.

For class 1: $$p(x|Class\,1)=\frac{1}{s_1\sqrt{2\pi}}e^{\frac{-(x-m_1)^2}{2s_1^2}}=\frac{1}{\sqrt{10}\sqrt{2\pi}}e^{\frac{-(x+4)^2}{2*10}}$$

For class 2: $$p(x|Class\,2)=\frac{1}{s_2\sqrt{2\pi}}e^{\frac{-(x-m_2)^2}{2s_2^2}}=\frac{1}{\sqrt{10}\sqrt{2\pi}}e^{\frac{-(x-5)^2}{2*10}}$$

In order to find he decision boundary, we have to solve the equation in the case $g_1(x) = g_2(x)$, which corresponds to P(Class1 | x) = P(Class2 | x).

Thus, we write the equation above using the Bayes formula for the posterior probability and solve it:

$$\frac{p(x|Class1)P(Class1)}{p(x)}=\frac{p(x|Class2)P(Class2)}{p(x)}$$

$$\frac{\frac{1}{\sqrt{10}\sqrt{2\pi}} e^{\frac{-(x+4)^2}{2*10}} 0.5}{p(x)} = \frac{\frac{1}{\sqrt{10}\sqrt{2\pi}} e^{\frac{-(x-5)^2}{2*10}} 0.5}{p(x)}$$

$$e^{\frac{-(x+4)^2}{2*10}} = e^{\frac{-(x-5)^2}{2*10}}$$

Notice that we can take the natural logarithm of both sides and simplify the exponential function:

$$\frac{-(x+4)^2}{2*10} = \frac{-(x-5)^2}{2*10}$$

$$-(x+4)^2 = -(x-5)^2$$

We have simplified the two discriminant functions a lot. On the left side you have $g_1(x)$ and on the right side of the equation you have $g_2(x)$ [in very simplified form].

To find the value of the decision criterion x* you just have to solve the equation.
The result is x* = 0.5

Note that to find the decision region for class 1, you have to determine the solution to $g_1(x) > g_2(x)$ – see the slides of the lecture and Assignment 7.3.

Thus, solving:

$$-(x+4)^2 > -(x-5)^2$$

you obtain
$x < 0.5$

which indicates the decision region for Class 1. All points that fall in that region will be classified as Class 1.
Since we have a dichotomizer classifier (only two classes), we can infer directly that the points that fall in $x > 0.5$ will be classified in Class 2.

Another easy way to determine the decision regions is to check on which side of the decision criterion the mean of the classes are. For example, $m_1=-4$ is on the left side of the decision criterion, thus the points $x<0.5$ are assigned to Class 1. (Note that this works only in the case of a linear dichotomizer).

b) *x_1 is in Class 2 (as $x_1 > x^*$, falls in the decision region of Class 2)*
c) *x_2 is in Class 1 (as $x_2 < x^*$, falls in the decision region of Class 1)*
d) *x_3 is in Class 2 (as $x_3 > x^*$, falls in the decision region of Class 2)*
e) *x_4 is in Class 1 (as $x_4 < x^*$, falls in the decision region of Class 1)*

Probability estimation – binary events

Assume that you have a rigged dice: out of six faces, it has one 1, one 2, one 3, two 4, and one 5. You roll the dice four times.

a) What is the probability of obtaining exactly two 4?
b) What is the probability of obtaining exactly four 2?

a)

The answer is 0.296

As the dice is rigged and we have two faces with the number 4, we know that the probability of getting a 4 is P(F) = 2/6. Consequently, the probability of not getting a 4 is P(¬F) = 4/6.

We apply the binomial distribution that allows to compute the probability of having exactly k successes (in our case a success is getting a 4, and to get it k=2 times on n=4 rolls of the dice)

$$\Pr(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$

$$P(F=2)=\binom{4}{2}[P(F)]^2[P(\neg F)]^2=\frac{4!}{2!2!}\left(\frac{2}{6}\right)^2\left(\frac{4}{6}\right)^2=6\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^2=0.296$$

b)

The answer is 0.00077

One out six faces of the dice has a 2. So the probability of getting a 2 is P(T) = 1/6. Consequently, the probability of not getting a 2 is P(¬T) = 5/6.

The probability of getting four times a 2, out of four dice rolls is:

$$P(T=4)=\binom{4}{4}[P(T)]^4[P(\neg T)]^0=\frac{4!}{4!0!}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^0=1\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^0=0.00077$$

Probability estimation – Bayes

You are standing in the box of the accused, on trial for murder. You want to scream your innocence from the rooftops!
You happen to live in the city where the crime was committed, but so do 100,000 people --- any of them could be the murderer!
Still, the evidence against you is overwhelming: DNA was found at the scene of the crime, and it was tested to match your DNA. This is the only evidence against you, but an expert witness is called to testify, and she testifies that the test is extremely reliable. The true positive rate of the test is 1 (that is, if the DNA really matches, the test always reports a match), and the false positive rate of a DNA test is one in 1,000,000 (that is, if the DNA is not actually a match, the test will still report a match once every 1,000,000 tests).

From the jury's perspective, what is the (approximate) probability that you are the murderer?

a. 999,999 / 1,000,000
b. 1 / 100,000
c. ~0.91
d. ~0.85
e. 0
f. ~0.99

The correct answer is **c) ~0.91**

Comments:
The jury wants to estimate the probability of you being the murdered, given that you were positive to a DNA match test. That is, they want to estimate the probability of an unknown event (you being the murderer), given that they collected evidence (a positive DNA match).
That is the posterior probability P(murdered | dnamatch).

If anyone in the city could be the murderer, it means that everyone has probability of 1 out 100,000 people to be the murderer. Thus, the prior probability of being a murdered is *P(murderer) = 1/100,000*
The prior probability of not being a murderer is
P(¬murderer) = 1 - P(murderer) = 99,999/100,000

They use a DNA test that is extremely reliable: the DNA test results in a positive match, given that it is done on the murderer, with probability 1 (certain event). This means that the probability of measuring a positive DNA match as evidence, given that we are testing the DNA of the murderer is 1. That is, the class-conditional probability is:
p(dnamatch | murderer) = 1

The DNA test can make errors, i.e. it can result positive also in cases in which the suspect is not the real murderer. It can result in a DNA match in 1 out of 1,000,000 cases, although the DNA is not of the murderer.
That is, there is a probability of 1/1,000,000 that it strikes a positive given that the person is not the murderer:

p(dnamatch | ¬murderer) = 1

We can use the Bayes formula to estimate the posterior probability.

$$P(murdered|dnamatch) = \frac{p(dnamatch|murderer)\,P(murderer)}{p(dnamatch)}$$

where the evidence p(dnamatch) is:

$$p(dnamatch) = p(dnamatch|murderer)\,P(murderer) + p(dnamatch|\neg murderer)\,P(\neg murderer)$$

We substitute the probabilities that we have estimated in the Bayes formula and obtain:

$$P(murdered|dnamatch) = \frac{1 * \frac{1}{100{,}000}}{1 * \frac{1}{100{,}000} + \frac{1}{1{,}000{,}000}, * \frac{99{,}999}{100{,}000}} = 0.91$$