SGN-41007 Pattern Recognition and Machine Learning
Exam 1.3.2019
Heikki Huttunen

- ▷ Use of calculator is allowed.
- ▷ Use of other materials is not allowed.
- ▷ The exam questions need not be returned after the exam.
- ▷ You may answer in English or Finnish.

1. Are the following statements true or false? No need to justify your answer, just T or F. Correct answer: 1 pts, wrong answer: $-\frac{1}{2}$ pts, no answer 0 pts.

   (a) The Receiver Operating Characteristics curve plots the probability of detection versus the probability of false alarm for all thresholds.

   (b) Random forest has a linear decision boundary.

   (c) Least squares estimator minimizes the squared distance between the data and the model.

   (d) The ReLU (rectified linear unit) activation function is defined as

   $$f(x) = \frac{1}{1 + \exp(-x)}.$$

   (e) Maxpooling returns the average within each disjoint block of neighboring samples.

   (f) Cross-validation is used for model accuracy evaluation.

2. *Maximum likelihood estimation.* Consider the model

   $$x[n] = A\exp(-n)\sin(\theta n) + w[n], \qquad n = 0, 1, \ldots, N-1,$$

   where $w[n] \sim \mathcal{N}(0, \sigma^2)$ and θ is a known real number. In other words, we assume that our measurement is a damped sinusoid at known frequency and phase and want to estimate the amplitude A. Derive the maximum likelihood estimator of A.

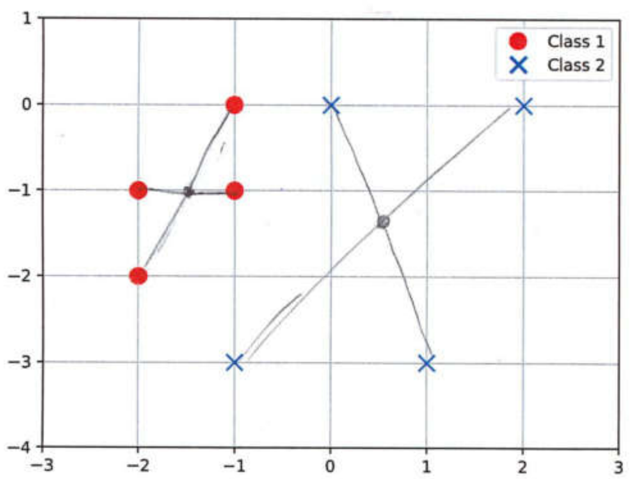

Figure 1: Training sample of question 3.

3. A dataset consists of two classes, containing four samples each. The samples are shown in Figure 1. The classes are linearly separable, and there are many linear decision boundaries that classify the training set with 100 % accuracy.

   (a) (4p) Find the Linear Discriminant Analysis (LDA) classifier for this data. The covariances of the classes are

$$C_1 = \frac{1}{3}\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \qquad C_2 = \frac{1}{3}\begin{pmatrix} 5 & 6 \\ 6 & 9 \end{pmatrix}.$$

   Present the decision rule for sample $\mathbf{x} \in \mathbb{R}^2$ in the following format:

$$\text{Class}(\mathbf{x}) = \begin{cases} 1, & \text{if } \boxed{\textit{something}} \\ 2, & \text{otherwise} \end{cases}$$

   (b) (2p) Find the threshold c at the center of the projected class means.

4. (6 pts) Consider the Keras model defined in Listing 1. Inputs are 96×96 color images from 10 categories.

   (a) Draw a diagram of the network.

   (b) Compute the number of parameters for each layer, and their total number over all layers.

5. A random forest classifier is trained on training data set and the `predict_proba` method is applied on the test data of five samples. The predictions and true labels are in Table 1.

   In the exercises we drew a receiver operating characteristic (ROC) curve. In this question, however, you are requested to draw a precision-recall-curve (PRC) instead, which is a curve drawn on precision-recall axes by sliding the detection threshold over all reasonable values. See also the supplementary material regarding the definitions of precision and recall.

   (a) Draw the precision-recall curve.

   (b) What is the Area Under the PR Curve?

	Prediction	**True label**
Sample 1	0.8	positive
Sample 2	0.3	positive
Sample 3	0.6	positive
Sample 4	0.4	negative
Sample 5	0.2	negative

Table 1: Results on test data for question 5.

Listing 1: A CNN model defined in Keras

```
model = Sequential()

w, h = 3, 3
sh = (3, 96, 96)

model.add(Convolution2D(30, w, h, input_shape=sh, border_mode='same'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Activation('relu'))

model.add(Convolution2D(40, w, h, border_mode='same'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Activation('relu'))

model.add(Convolution2D(50, w, h, border_mode='same'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Activation('relu'))

model.add(Convolution2D(60, w, h, border_mode='same'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Activation('relu'))

model.add(Flatten())
model.add(Dense(10, activation = 'softmax'))
```

Related Wikipedia pages

Inversion of 2 × 2 matrices [edit]

The *cofactor equation* listed above yields the following result for 2 × 2 matrices. Inversion of these matrices can be done as follows:[6]

$$\mathbf{A}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

ROC space [edit]

The contingency table can derive several evaluation "metrics" (see infobox). To draw a ROC curve, only the true positive rate (TPR) and false positive rate (FPR) are needed (as functions of some classifier parameter). The TPR defines how many correct positive results occur among all positive samples available during the test. FPR, on the other hand, defines how many incorrect positive results occur among all negative samples available during the test.

A ROC space is defined by FPR and TPR as *x* and *y* axes respectively, which depicts relative trade-offs between true positive (benefits) and false positive (costs). Since TPR is equivalent to sensitivity and FPR is equal to 1 − specificity, the ROC graph is sometimes called the sensitivity vs (1 − specificity) plot. Each prediction result or instance of a confusion matrix represents one point in the ROC space.

		True condition				
	Total population	Condition positive	Condition negative	Prevalence $= \frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$	Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$	
Predicted condition	Predicted condition positive	**True positive**, Power	**False positive**, Type I error	Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$	
	Predicted condition negative	**False negative**, Type II error	**True negative**	False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$	
		True positive rate (TPR), Recall, Sensitivity, probability of detection $= \frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm $= \frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$	Positive likelihood ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR−}}$	F_1 score = $\frac{1}{\frac{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}{2}}$
		False negative rate (FNR), Miss rate $= \frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$	Negative likelihood ratio (LR−) $= \frac{\text{FNR}}{\text{TNR}}$		