Exercise 9 : Exam preparation

Exercise 9-1 : Bayes optimal classifier

We have a classification problem with two classes "+" and "−" and three trained classifiers h_1, h_2, and h_3 with the following probabilities of the classifiers, given the training data D :

$$\begin{aligned}\Pr(h_1|D) &= 0.6\\ \Pr(h_2|D) &= 0.3\\ \Pr(h_3|D) &= 0.1\end{aligned}$$

For the five test instances o_1, o_2, o_3, , o_4, o_5 the classifiers give the following class probabilities :

$$\begin{aligned}o_1 : \Pr(+|h_1) &= 0.4\\ \Pr(+|h_2) &= 0.7\\ \Pr(+|h_3) &= 0.5\\ o_2 : \Pr(+|h_1) &= 0.1\\ \Pr(+|h_2) &= 0.5\\ \Pr(+|h_3) &= 0.8\\ o_3 : \Pr(+|h_1) &= 0.0\\ \Pr(+|h_2) &= 0.4\\ \Pr(+|h_3) &= 0.6\\ o_4 : \Pr(+|h_1) &= 0.2\\ \Pr(+|h_2) &= 0.7\\ \Pr(+|h_3) &= 0.5\\ o_5 : \Pr(+|h_1) &= 0.5\\ \Pr(+|h_2) &= 0.1\\ \Pr(+|h_3) &= 0.2\end{aligned}$$

We combine the three classifiers to get a Bayes optimal classifier. Evaluate the correctness of the probability calculations below for test instances o_1, o_2, o_3 we will get from this Bayes optimal classifier.

(a) $o_1 : \Pr(+|\text{Bayes optimal}) = \Pr(-|\text{Bayes optimal})$

(b) $o_2 : \Pr(+|\text{Bayes optimal}) > 0.5$

(c) $o_3 : \Pr(-|\text{Bayes optimal}) > 0.5$

(d) $o_4 : \Pr(-|\text{Bayes optimal}) < 0.5$

(e) $o_5 : \Pr(-|\text{Bayes optimal}) < 0.5$

Exercise 9-2 : Naive Bayes classifier

Consider a classification problem where the goal is to map two features $X_1 \in \{A, B, C\}$ and $X_2 \in \{D, E, F, G\}$ to class labels $Y \in \{-1, +1\}$. We train a naive Bayes classifier on the data set below

ID	X_1	X_2	Y
1	C	F	+1
2	A	E	-1
3	A	G	+1
4	B	G	-1
5	C	F	-1
6	B	D	+1
7	B	E	+1
8	C	E	-1

Evaluate the following statements for this naive Bayes classifier.

(a) For test query (C, D), it predicts higher probability for class $+1$ than class -1.

(b) For test query (A, D), it predicts the class label as $+1$ with probability lower than 0.95.

(c) For test query (B, F), it predicts the class label as $+1$ with probability higher than 0.5.

(d) For test query (A, F), it predicts the class label as -1 with probability higher than 0.25.

(e) For test query (C, G), it predicts the class label as $+1$ with probability higher than 0.25.

Exercise 9-3 : Precision-recall

The true class $f(o)$ of 10 test objects o and the corresponding predictions of some classifier $h(o)$ are given in the table below.

o	true class ($f(o)$)	prediction ($h(o)$)
o_1	A	A
o_2	A	B
o_3	A	A
o_4	A	A
o_5	A	B
o_6	B	A
o_7	B	B
o_8	B	B
o_9	B	B
o_{10}	B	B

Evaluate statements below about the class-specific recall and precision scores for this classifier.

(a) Recall for class A is greater than recall for class B.

(b) Precision for class A is greater than precision for class B.

(c) Recall for class A is greater than precision for class A.

(d) Recall for class B is greater than precision for class B.

(e) Recall for class A is greater than precision for class B.