IN3310 2025 UiO - Exercise Week 8

1 Average Precision in Extreme Cases

a

Suppose the set used for evaluation consists of 11 samples, of which 3 are actual positives. Those 3 have a *higher* prediction score for the positive class than the other 8 samples. What is the average precision?

b

Suppose the set used for evaluation consists of 11 samples, of which 3 are actual positives. Those 3 have a *lower* prediction score for the positive class than the other 8 samples. What is the average precision?

c

Suppose the set used for evaluation consists of N samples, of which R are actual positives, and those R samples have *lower* prediction scores for the positive class than the other $N - R$ samples. What is the average precision?

Hint: The answer can be an expression that contains a summation and depends on R and N. You will (likely) not be able to write it without a summation.

2 Average precision of random predictor

Suppose the set used for evaluation consists of N samples, of which R are actual positives. A predictor is used to calculate prediction scores for the positive class for each sample.

a

Assume the ranking of the scores evenly distributes the R actual positive samples. More specifically, after sorting the samples by their scores in descending order and letting 1 be the first index (the sample with the highest score), assume that the first actual positive sample has index $\frac{N}{R}$ (and that this is an integer), the second actual positive sample has index $2\frac{N}{R}$, and so on, making the l^{th} actual positive sample appear at index $l\frac{N}{R}$ and the last actual positive sample appears at index $R\frac{N}{R} = R$. What is the average precision?

Hint 1: Calculate the precision at $k(P@k)$ for every index of an actual positive sample, i.e., for $k = l\frac{N}{R}$ with l = 1, 2, . . . , R.

Hint 2: It is possible to write the average precision without any summation.

b

Again, assume that the ranking of the scores evenly distributes the R actual positive samples, but this time, the order is shifted such that the first actual positive sample has index 1. More specifically, after sorting the samples by their scores in descending order and letting 1 be the first index (the sample with highest score), the first actual positive sample has index 1, the second actual positive sample has index $1 + \frac{N}{R}$ (and this is assumed to be an integer), and so on, making the l^{th} actual positive sample appear at index $1 + (l-1)\frac{N}{R}$ and the last actual positive sample appear at index $1 + (R-1)\frac{N}{R}$. What is the average precision?

Hint: The answer can be an expression that contains a summation and depends on R and N. You will (likely) not be able to write it without a summation.

c

Will the average precision in subtask **b** be lower or higher than the average precision in subtask **a**?

3 Average Precision and Accuracy of a Linear Classifier

Consider a linear classifier $s(x) = wx + b$ classifying samples as positives if $s(x) > 0$ and otherwise as negatives. w and b are trainable parameters.

a

Which of the trainable parameters will impact the resulting average precision calculated using the prediction scores for the positive class, $s(x)$? Please explain.

b

Which of the trainable parameters will impact the resulting accuracy? Please explain.

4 Area Under the Receiver Operating Characteristic Curve (AUROC)

a

Assume that we observe that a model obtains an AUROC of 1 when using a particular evaluation set. What does that imply for the ranking of the prediction scores for the positive class in that evaluation set?

b

Suppose the set used for evaluation consists of N samples, of which 2 are actual positives. One of these 2 actual positives has a higher prediction score for the positive class than the other $N1$ samples. The other of these 2 actual positives have a prediction score for the positive class that is greater than 80% of the prediction scores for the positive class for actual negative samples (and there are no ties, so 20% of the actual negative samples have a higher score than 1 of the 2 actual positives). What is the AUROC?

c

Suppose the set used for evaluation consists of N samples, of which 4 are actual positives. One of these 4 actual positives has a higher prediction score for the positive class than the other $N1$ samples. Two of the other of these 4 actual positives have a prediction score for the positive class that is greater than 80% of the prediction scores for the positive class for actual negative samples (and there are no ties, so 20% of the actual negative samples have a higher score than these 2 of the 4 actual positives). The last of these 4 actual positives has a prediction score for the positive class that is greater than 60% of the prediction scores for the positive class for actual negative samples (and there are no ties, so 40% of the actual negative samples have a higher score than 1 of the 4 actual positives). What is the AUROC?

d

Suppose the evaluation set and scores are the same as described in subtask **c**. In order to obtain a dichotomous classifier, the score is thresholded such that 3 of the 4 actual positives and 20% of the actual negatives are classified as positives, while the last actual positive and 80% of the actual negatives are classified as negatives. What is the AUROC of this binary classifier?