Exam Preparation
Machine Learning S. 5 Bachelor WS21/22

Jonas Weßner

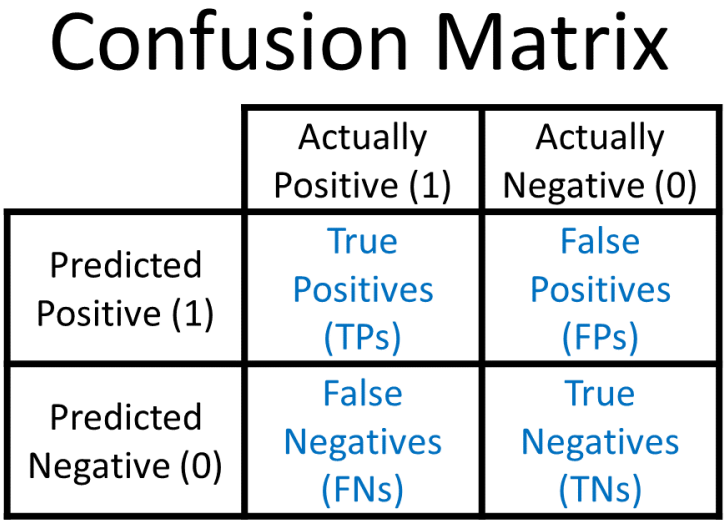
February 4, 2022

Contents

1 Metrics for evaluating predictions

The following metrics can be used to analyze the quality of a *classification model*.

1.1 Confusion Matrix



1.2 Accuracy

Accuracy answers the question " *What is the probability that a prediciton is correct?*".

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

It is only good, if the real distribution of positive and negatives in the data is close to symmetric.

1.3 Precision

Precision answers the question "*If we classify something as positive, how probable is it that it is actually positive?*".

$$Precision = \frac{TP}{TP + FP}$$

1.4 Recall

Recall a.k.a. sensitivity answers the question "*If a sample is positive, what is the probability we also label it as positive?*".

$$Recall = \frac{TP}{TP + FN}$$

1.5 F1 Score

The F1-score divides the true positives by the sum of the true positives and the mean of the false positives and false negatives. This a high F1-score requires the model to make not few false predictions in either direction. Therefore F1-score is better than accuracy if the real distribution of positive and negative values in the dataset is uneven.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = \frac{TP}{TP + \frac{1}{2} \cdot (FP + FN)}$$

Describe here: Size of inputs and outputs in MLP and convolutional NN calculated from image size and the number of output classes.

11 Activation functions

11.1 Softmax

11.2 Sigmoid

11.3 RELU

12 Solving non-linear problems with NNs

Use example of logical function XOR here.

13 K-means

14 Gradient Descent

15 Hyperparameters of ML models

15.1 Learning Rate

15.2 Epochs

15.3 Regularization

15.4 Batch Size

15.5 Convolution Kernel size

15.6 Max-Pooling

16 Logistic Regression and Cross Entropy

17 Linear Regression and Normal Equation

18 Decision Trees

19 K-nearest Neighbors