**Aantekeningen recap 2**

Decision trees

Supervised classification is with discrete class labels as output. In regression you ise Continuous (real number).

If you have 50 questions, decision tree is not good. Questions should be limited, not more then 20 questions.

**Evaluation metrics**

Accuracy is not the only measure that matters. How well is the model doing?

Can you make the quality measurable?

Confusion matrix, only use letters:

TP, FN, FP, TN. Always have it in the same order.

Accuracy = (TP + TN) / (TP + TN + FP + FN)

You want to avoid FN

Recall: how many did we correctly diagnose as sick? TP / (TP + FN). Horizontal line.

Precision (avoiding False Positives): How many were actually spam? TP / (TP + FP) Vertical line.

F1 score: Combining Precision and Recall.

Calculate precision, calculate recall, you do that 2 times and then devide it by the two together.

F1 = 2 x P x R / (P+R)

Recall and Precision are most important.

**Type of numerical data**

* Discrete, should be countable
* Continiuous, measurements, ratios

Continuous labelling

You have some observe points.

Regression is normally with continuous variables.

ANN, CNN, NLP, Reinforcement learning modellen heb je nodig dit semester