SONAT

Machine learning quiz

Created by: Tommy Odland

Instructions

Please read the entire problem sheet before starting.
If you get stuck, move to a different problem.

1. **Multiple choice**

   (a) Linear models are linear. But with respect to what exactly?
   ◯ Input variables x_i ◯ Features ϕ_i ◯ The function space

   (b) Which loss function in a linear model deals well with outliers?
   ◯ ℓ_1 loss (manhattan) ◯ ℓ_2 loss (euclidean) ◯ ℓ_∞ loss (minmax)

   (c) What concept(s) do we associate with logistic regression?
   ☐ AI ☐ GLM ☐ Binary classification

   (d) Which abstract class(es) are in the `sklearn` API?
   ☐ Predictor ☐ Regressor ☐ Transformer

   (e) Which two of these are sensible machine learning use cases?
   ☐ Deciding if sending emails leads to more sales
   ☐ Predicting student grades in absence of exams due to coronavirus
   ☐ Predicting which horse to bet money on in horse races
   ☐ Deciding how to plan infrastructure in a city
   ☐ Sorting an array
   ☐ Finding the best way to transmit signals through a large network
   ☐ Modeling the spread of disease
   ☐ Predicting incoming calls in a call center

   (f) What is the primary advantage of LASSO (ℓ_1 regularization)?
   ◯ Rotational invariance ◯ Smaller weights ◯ Sparse weights

   (g) How do we deal with nominal data (e.g. colors) in a linear model?
   ◯ Remove it ◯ One-hot-encoding ◯ Use imputation

2. **Write a sentence answering these questions**

(a) What are the advantages and disadvantages of linear models compared to more complex models? (for instance neural networks)

(b) What are common ways of dealing with missing data entries?

(c) What is regularization, and why is it often a good idea?

(d) What is boosting?

(e) What is the purpose of the test data set? How is it used?

(f) Name three things to think about before putting ML in production.