Final Exam, Machine Learning (MIRI) June 16th, 2021

Instructions

- You have **2h** to solve the exam.
- Please **note your seat for covid-tracking purposes in all your pages** together with your full name.
- You **can** use:
  - Any paper notes and books you bring to the exam
  - A calculator
  - A PDF viewer on your laptop/tablet for lecture slides or other previously downloaded material
- You **cannot** use:
  - Any connectivity to wifi or network at all. Make sure you download all material before the exam.
  - Any software on your laptop, including python/R/matlab or numeric solvers.

  If in doubt, ask for permission before. Any violation of this rule will be considered cheating.
- Be concise but clear. We may penalize answers that are unnecessarily lengthy.
- Good luck!

All questions have equal weight. Justify all your answers except Question 5.

Question 1

Ordinal regression refers to a type of supervised learning problem where (discrete) taget labels show a natural ordering. For example, classifying wines in a 1-10 scale, or predicting customer satisfaction into one of `excellent`, `good`, `average`, or `bad`.

Given the type of supervised learning problem that you have come accross during our course, what modifications or extensions can you think of that could solve the ordinal regression problem? You can mention how to alter learning algorithms, or cost functions, or anything you can think of.

Question 2

A data analyst has received a (small) set of expensive, labelled data and wants to build a good predictive classification model. She comes up with the following protocol:

1. Split all available data into training, validation and test sets using 50/25/25 proportions at random
2. Train SVM model with default parameters and polynomial kernel of degree 3 on the training set
3. Train and optimize Random Forest, optimizing its hyper-parameters using OOB on the training set
4. Choose best of SVM, and RF models using error on the validation set
5. Estimate true error of selected model using test set

Please criticize (in a constructive manner) this solution.

Question 3

During our last class, we saw that if k_1 and k_2 are *kernel* functions on $\mathbb{R}^d$, then

$$k_3(u, v) := k_1(u, v) + k_2(u, v)$$

is also a kernel (with $u, v \in \mathbb{R}^d$). Please describe the feature map associated with this new kernel k_3.

Question 4

Explain in your own words the difference between the **posterior** and the **predictive** distribution in Bayesian learning.

Question 5

Please mark whether the following statements are **true** or **false**; the score for this question is given by the formula $2 \times \frac{\text{nr. of correct} - \text{nr. of incorrect questions}}{15}$ when positive, otherwise it is 0.

- ☐ Training error is always lower than test error False
- ☑ Lasso and ridge regression both help in reducing overfitting True
- ☐ Lasso regression is preferable to ridge regression because it produces sparse models False, depends sparse model is not always good
- ☐ The activation functions of output neurons of a neural network are determined mainly by the nature of the target variable one wants to predict
- ☐ Linear regression assumes Gaussian input variables False, assumes Gaussian response
- ☐ Naive Bayes assumes Gaussian input variables FALSE: a particular case, Gaussian NB does, but in general NB can be applied to any combination of input variable distributions
- ☑ Bayes formula is used in Bayesian learning to obtain posterior distributions
- ☐ It is not possible to train a neural network for both regression and classification at the same time FALSE:all we need to do is have a differentiable error function combining both regression error and classification error; a simple sum may do
- ☑ Bigger training sets help to reduce overfitting with more and more data overfitting becomes harder
- ☐ It is impossible to evaluate the quality of a clustering result because we never have the ground truth
- ☑ Cross-validation is a resampling method used to select a good model
- ☑ Gaussian Naive Bayes assumes Gaussian input variables
- ☐ Backprop is an algorithm used in neural network learning to obtain partial derivatives of an error function with respect to its weights
- ☑ The negative log-likelihood can always be used as an error function in supervised learning
- ☑ The EM algorithm is particularly suited to learn probabilistic models with partially observed data