Syllabus

Contents

In this course, we will cover evaluation metrics for classification and regression, feature engineering, feature selection, feature importances, ensembles, and regularization.

Course learning outcomes

By the end of the course, students are expected to be able to

- build, debug, appropriately evaluate, and refine supervised machine learning models
- appropriate choose the right evaluation metric for classification and regression problems
- broadly describe and carry out preliminary feature engineering
- explain different feature selection methods and carry out feature selection
- interpret machine learning models in terms of feature importances
- explain and carry out L1- and L2-regularization

Conda environment setup

To set up the necessary packages for running the labs and lecture material, download the environment file from the student repo to your computer (hit "Raw" and then `Ctrl` + `s` to save it, or copy paste the content). Then create a virtual environment by using `conda` with the environment file you just downloaded:

```
conda env create --file environment.yaml
```

This will setup Python with the correct versions of all required packages.

> **Note**
>
> If you don't remember how to use a conda environment with JupyterLab, review this section of Lec 8 in 521. In essence, you should only need to run the following command from the environment where you have JupyterLab installed (e.g. `base` or `jl`) if you haven't already done so.
>
> ```
> conda install nb_conda_kernels
> ```
>
> Next time you open JupyterLab, you should be able to select the "573" as your notebook kernel.

Assessments

- The first few questions in each lab will have all their tests visible for you (worksheet-style). The reason for this is to reduce the number of separate documents and submissions that you have to keep track of.
- The weight for the assessments can be seen below; due dates can be found on Gradescope.

Assessment	Weight
Lab 1	12%
Lab 2	12%
Quiz 1	25%
Lab 3	12%
Lab 4	12%
Quiz 2	25%
iClicker	2%

Lectures

Format

This course will be run in person. We will meet three times every week: twice for lectures and once for the lab. You can refer to the Calendar for lecture and lab times and locations. Lectures of this course will be a combination of slides, discussions, a few pre-recorded videos, and class activities. The slides will highlight the most important concept from the lectures notes, and the rest of the notebook are considered required reading after lectures. It's optional but recommended to download the appropriate datasets provided below and put them under your local `lectures/data` directory, and run the lecture Jupyter notebooks on your own and experiment with the code.

Lecture schedule

This course occurs during **Block 3** in the 2022/23 school year.

Lecture	Topic	Assigned videos	Resources and optional readings
1	Evaluation metrics for classification, class imbalance	📹 • post-lecture: 9.4	• ROC animation
2	Evaluation metrics for regression	📹 • pre-lecture: 10.1	
3	Feature engineering	None	• spaCy • The Learning Behind Gmail Priority Inbox • Google n-gram viewer
4	Feature importances and selection	None	
5	Loss functions and regularization	None	• Mike's video on loss functions
6	More regularization	None	
7	Ensembles	📹 • 11.1	
8	Model interpretability and SHAP	📹 • 12.1	• SHAP • Interpretable Machine Learning

Datasets

Here is the list of Kaggle datasets we'll use in this class.

- Credit Card Fraud Detection
- Housing Prices dataset
- The adult census dataset
- Spotify Song Attributes

Reference Material

▶ Click to expand!

Policies

Please see the general MDS policies.

COVID-19 safety

Read the UBC COVID-19 Campus Rules for the latest updates of what is expected of you in terms of COVID-safety.