Syllabus

NEU 365P/385P - Spring 2024 - Programming and Data Analysis for Modern Neuroscience

Python intro

1. ▶ Jan 16 T - Course intro and python intro
    - You will get a brief overview of the course.
    - You will be able to use `conda` to manage python environments.
    - You will be able to use `conda` and `pip` to manage python packages within each environment.
    - You will be able to run python code in a Jupyter notebook.
    - You will understand some basic python code:
        - Variables
        - Types
        - Basic operations
        - Logical comparisons
        - Comments
        - String formatting
2. ▶ Jan 18 R - Python basics
    - You will understand some basic python code:
        - `if` code blocks
        - Nested code blocks
        - `list` and index/slice
        - `dict` (key,value) pairs
        - `for` and `while` loops
        - Functions (optional named and default arguments)
        - Assignment vs mutation
3. ▶ Jan 23 T - Classes and modules
    - You will understand how to compartmentalize python code beyond simple functions:
        - `class` code blocks
        - Class `__init__` method
        - Class instance (`self`) vs class template
        - Class inheritance
        - Modules
    - You will have heard from me that shoehorning your code into classes is often *unnecessary overcomplication*, whereas modules are almost always a good idea for anything larger than a short script.

Working with data

4. ▶ Jan 25 R - N-dimensional arrays and basic plots
    - You will appreciate that many types of data can be represented as N-dimensional arrays.
    - You will understand how to work with `numpy` N-dimensional arrays:
        - Array initialization (e.g., `zeros`, `ones`, `random`) and `shape`
        - Element-wise array math

        - Index and slice
        - Logical masks
        - Reductions (e.g., `min`, `max`, `mean`, `var`)
        - Broadcasting
    - You will appreciate that `numpy` can be *much much* faster than raw python.
    - You will appreciate that without `numpy` we would not use python for most data analysis.
    - You will be able to visualize data with simple plots using `matplotlib`.
5. ▶ Jan 30 T - Tabular data and basic plots
    - You will be able to to work with tabular data sets using `pandas`:
        - Convert between `pandas` dataframes and `numpy` arrays.
        - Read/Write `pandas` dataframes from/to `*.csv` or Excel files.
        - Index and slice like numpy (e.g., `iloc`) or by name (e.g., `loc`)
        - Logical masks
        - Missing values
        - Column-wise reductions (e.g., `sum`, `mean`)
        - Group data (e.g., `groupby`)
        - Simple plots (e.g., `plot`, `plot.bar`)
        - Correlations
    - You will be able to use `seaborn` to create some nice looking plots from a `pandas` dataframe.
    - You will be able to use `hvplot` to create some nice looking plots from a `pandas` dataframe.
    - You will appreciate how useful `pandas` is for exploratory data analysis.

Probability and random variation

6. ▶ Feb 01 R - Probability distributions
    - You will understand the difference between a probability and a probability density.
    - You will be able to compute some common descriptive statistics (e.g., mean, variance).
    - You will understand how some basic probability distributions relate to particular types of random behavior:
        - **Normal**: random fluctuations (e.g., white noise)
        - **Exponential**: random intervals between events ocurring at a constant average rate (e.g., time between spikes for a spiking neuron)
        - **Poisson**: random number of events within an interval for events ocurring at a constant average rate (e.g., number of spikes in a second for a spiking neuron)
        - **Binomial**: random number of successes for some number of trials all with the same probability of success (e.g., number of times subject recieved reward out of total number of trials)
    - You will be able to visualize how well a probability distribution explains data.
    - You will be able to use a probability distribution to make predictions.

Resampling

7. ▶ Feb 06 T - Bootstrap confidence interval and permutation test
    - You will understand the difference between a population distribution and a sample.
    - You will appreciate that statistics for different samples are likely to vary.
    - You will understand the concept of a sampling distribution.
    - You will understand the concept of a confidence interval.

- You will be able to compute a confidence interval using bootstrapping.
- You will be able to test the hypothesis that two samples come from the same population distribution using a permutation test.
- You will appreciate how the Central Limit Theorem explains why normal-ish distributions are frequently observed in biological measurements.

Model optimization

8. ▶ Feb 08 R - Curve fitting and maximum likelihood estimation (MLE)
    - You will be able to fit a function to data by minimizing the residuals.
    - You will be able to fit an arbitrary probability distribution to data by maximizing the loglikelihood.
    - You will understand the concept of gradient descent minimization.
    - You will appreciate the difference between local and global optimization.

Review

9. Feb 13 T - Review

Linear model

10. ▶ Feb 15 R - Linear regression
    - You will be able to fit a line to X vs. Y data.
    - You will be able to fit a (hyper-)plane to {X0, X1, X2, ...} vs. Y data.
    - You will be able to predict the Y value for new {X0, X1, X2, ...} values.
    - You will be able to compute the mean squared error (MSE) and R^2 value for your fit.
    - You will be able to compute confidence intervals for all model parameters and visualize a confidence envelope for your fit.
    - You will appreciate why the residuals should be normally distributed.
    - You will appreciate why data points with high leverage can greatly influence your fit.
    - You will understand under what conditions you may want to standardize your features {X0, X1, X2, ...}.
    - You will understand that regression involves modeling a relation between feature variables {X0, X1, X2, ...} and a target variable Y.
    - You will appreciate that it is straightforward to understand the meaning of the parameters in a linear regression.
    - You will appreciate that the existence of a mathematically computable solution makes linear regression extremely fast.

Nonlinear model

11. ▶ Feb 20 T - Polynomial and k-nearest neighbors (KNN) regression
    - You will be able to fit a polynomial to {X0, X1, X2, ...} vs. Y data.
    - You will be able to fit nonlinear {X0, X1, X2, ...} vs. Y data with a KNN model.
    - You will be able to predict the Y value for new {X0, X1, X2, ...} values.
    - You will understand how polynomial regression can be recast as a simple linear regression.
    - You will appreciate that although a KNN model can be used to explain or predict lots of arbitrary nonlinear relations, it is less obvious what the model means.

Model selection

12. ▶ Feb 22 R - Train/Test, bias/variance, and cross validation
    - You will be able to split your dataset up into training and testing sets.
    - You will understand the difference between training error and testing error.
    - You will appreciate that often the best model is the one that will generalize best to new data (i.e., has the lowest testing error, not the lowest training error).
    - You will understand the concept of the "bias vs. variance" tradeoff.
    - You will be able to perform K-fold cross validation.

Regularization

13. ▶ Feb 27 T - Ridge and lasso regression
    - You will appreciate how correlations can influence a linear regression.
    - You will be able to perform ridge and lasso regression.
    - You will appreciate how regularization can prevent poorly constrained model parameters from exploding.
    - You will appreciate how lasso regularization can identify model parameters with little to no impact.
    - You understand how to choose (tune) the regularization hyperparameter.

Generalized linear model (GLM)

14. ▶ Feb 29 R - Poisson and logistic regression
    - You will gain a conceptual understanding for a generalized linear model.
    - You will appreciate why a GLM may be a better choice than a simple linear model for neural spiking data.
    - You will use a GLM (poisson regression) to predict a neuron's spiking in response to a stimulus.
    - You will see how the choice of noise distribution in a GLM can be used for binary classification.
    - You will use a GLM (logistic regression) to predict a mouse's left vs. right choice from its neural activity.

Classification

15. ▶ Mar 05 T - Confusion matrix, ROC curve, support vector machine (SVM)
    - You will understand that classification invovles modeling the categorical grouping of data.
    - You will be able to use a logistic regression binary classifier.
    - You will be able to use your classifer to predict the class to which data belongs.
    - You will be able to compute the accuracy of your classifier given data with known class labels.
    - You will be able to use your classifer to get the probability of each possible class.
    - You will be able to compute cross validated predictions, accuracy, and probabilities.
    - You will be able to generate a confusion matrix for your classifier.
    - You will be able to generate a ROC curve for your classifier.
    - You will gain a conceptual understanding for classification with a support vector machine.
    - You will be able to use a SVM classifier to separate data with linear boundaries.
    - You will appreciated at the conceptual level that SVM can achieve complex nonlinear boundaries by projecting the data into higher dimensions.
    - You will be able to use a SVM classifier to separate data with nonlinear boundaries.

Ensemble model

!!! Skipped in favor of more time on GLMs and Classification

16. ▶ Mar 07 R - Random forest classifier and extreme gradient boosting (XGBoost)
    - You will gain a conceptual understanding of a decision tree.
    - You will gain a conceptual understanding of a random forest ensemble of decision trees.
    - You will be able to use a random forest classifier.
    - You will understand the concept of boosting.
    - You will be able to use a XGBoost classifier.

⛱️ Spring break

- Mar 12 T - SPRING BREAK
- Mar 14 R - SPRING BREAK

Time series (i.e., sequences)

17. ▶ Mar 19 T - Time series and convolution
    - You will appreciate that data points in sequences are correlated (unless pure noise) as opposed to independent random variables.
    - You will be able to compute the autocorrelation of a sequence.
    - You will be appreciate how undersampling can introduces aliasing artifacts in a sequence.
    - You will be able to visualize the frequency power spectrum of a 1-D sequence.
    - You will be able to visualize the frequency spectrrogram of a 1-D sequence.
    - You will understand why convolution describes a systems output based on its impulse response.
    - You will be able to convolve two 1-D sequences.
    - You will appreciate how convolution can be used to filter a sequence.
    - You will be able to apply lowpass, highpass and bandpass finite impulse response (FIR) filters to a 1-D sequence.
    - You will be able to properly downsample a 1-D sequence without introducing alisasing artifacts.
    - You will be able to convolve two 2-D sequences (e.g., images).
    - You will appreciate that convolution can be used to highlight features in an image.
    - You will appreciate that the joint probability distribution resulting from adding two random variables is the convolution of their individual probability distributions.

Simulating a neuron

18. ▶ Mar 21 R - Leaky integrate and fire (LIF) neuron
    - You will appreciate how a cell membrane can be described by a RC circuit.
    - You will understand the concept of the LIF neuron model.
    - You will be able to simulate a LIF neuron.
    - You will be able to plot spike rasters.
19. ▶ Mar 26 T - LIF neuron with synaptic input
    - You will be able to simulate stochastic synaptic input to a LIF neuron.
    - You will appreciate how convulation can be used to integrate synaptic inputs.

Hidden Markov model

20. ▶ Mar 28 R - Hidden Markov model (HMM) for an ion channel
    - You will understand the concept of a hidden Markov model.
    - You will use an HMM to model current flowing through a single ion channel.
    - Given an HMM, you will be able to compute the most likely state trajectory for a data sequence.
    - You will appreciate how an HMM uses the full sequence to inform the model.
    - You will use the Bayesian information criterion (BIC) to choose the best model out of several possibilities.
21. ▶ Apr 02 T - HMM for a DNA sequence
    - You will use an HMM to predict exons and introns in a nucleotide sequence.

Clustering

22. ▶ Apr 04 R - K-means, gaussian mixture model (GMM), DBSCAN, etc.
    - You will appreciate the difference between classification and clustering (i.e., no labels to train on).
    - You will understand and be able to use several different clustering algorithms to segregate data.
    - You will appreciate that each clustering algorithm has its own pros and cons.
    - You will be able to use several different empirical metrics to choose an optimal clustering model (e.g., number of clusters).
    - You will use the Bayesian information criterion (BIC) to choose the optimal number of clusters for a GMM.

Dimensionality reduction

23. ▶ Apr 09 T - Principal component analysis (PCA)
    - You will visualize the process of changing your perspective to align with the variance in the data.
    - You will visualize the effects of projecting the data onto a smaller number of principal components.
    - You will be able to interpret the principal components as axes in the original data space.
    - You will be able to quantify the amount of variance explained by any given number of principal components.
    - You will understand how images can be represented as points in a high dimensional space.
    - You will be able to apply PCA to images.
    - You will see how PCA can be used as a filter to remove noise.
24. ▶ Apr 11 R - PCA for EEG and RNAseq data (or UMAP or t-SNE)
    - You will apply PCA to EEG time series.
    - You will be able to cluster time series and visualize the clustering in a low number of PCs.
    - You will appreciate how clustering of time series could be beneficial for interpreting experimental data.
    - You will walk thorugh an example of clustering in reduced dimensions for single cell RNAseq data.
    - You will appreciate the importance of being able to think critically about your data.

Neural network

25. ▶ Apr 16 T - Feedforward neural network (FNN)
    - You will understand the basic concept of a neural network as a universal function generator.
    - You will understand how the input and output layers of a neural network depend on the data and desired computation.
    - You will understand the concept of how a neural network is trained.
    - You will be able to implement basic feed-forward neural networks for regression and classification in Python.
    - You will appreciate that neural networks are not always the best choice.
26. ▶ Apr 18 R - Convolutional neural network (CNN)
    - You will be able to implement neural networks using PyTorch.
    - You will understand the basic concept of a CNN.
    - You will apply a CNN to decipher grating orientations based on images of gratings.
27. ▶ Apr 23 T - Recurrent neural network (RNN)
    - You will understand the basic concept of a RNN.
    - Long/Short term memory (LSTM) neural network

Exam

28. Apr 25 R - Exam