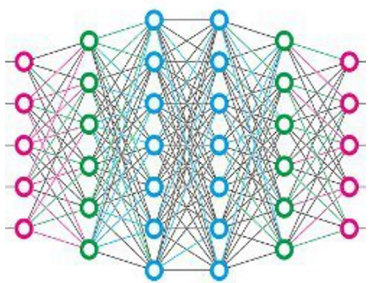
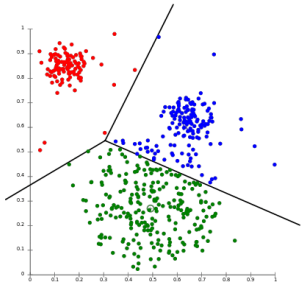
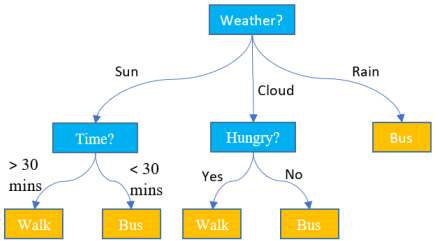
Welcome to CS 445
Introduction to Machine Learning
Model Evaluation, Selection, and Validation

Instructor: Dr. Kevin Molloy







Announcements

- Quiz 2 on Canvas due Wednesday by 11:59 pm
- PA 1 is due next Friday (Sept 18th).
- Video is posted (link in Canvas under modules, week 2) that discusses Decision Trees for regression. Watch and we will briefly discuss next class.

Learning Objectives for Today

- Define and discuss the differences between model evaluation and model selection
- Define the term hyper-parameter and make the distinction between it and a parameter
- Utilize training, validation, and test sets to design experiments to better characterize your model's performance
- Define and utilize k-fold cross validation

Plan for Today

- Complete Lab questions 1 and 2 (6 to 8 minutes)
- Discussion

- Complete Lab questions 3 to 7(skip question #8, it is repeated)
- Discussion

- Complete Lab questions 11
- Discussion

- General discussion
- **Submit** completed PDF to Canvas

Model Selection

The task of finding the model that maximizes the performance of **learning** task is called **model selection**. This involves tuning **hyper-parameters**

What are the hyper-parameters?

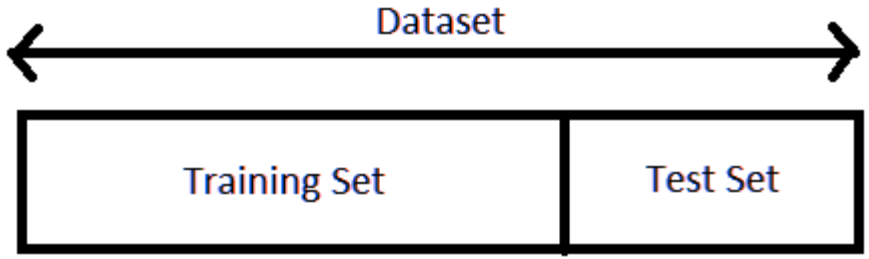
- Parameters NOT directly learned from the training data
- Set before the induction(learning) process begins

Plan for Today

- Complete Lab questions 1 and 2 (6 to 8 minutes)
- Discussion

Model Selection

Goal is to minimize **generalization error**.



Separate data into two groups: training and test. For each hyper-parameter setting:

- Build the model using the training set
- Evaluate the model using the test set

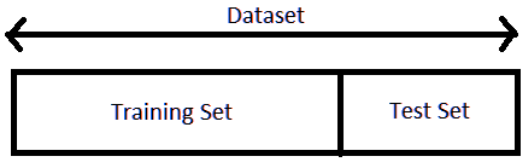
Next Steps

- Complete Lab questions 3 – 9 (15 minutes)
- Discussion

Dataset Purposes

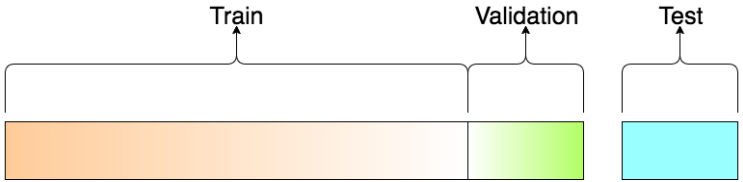
Goal is to minimize **generalization error**. However, we are using the test set for two purposes:

- Model selection (best hyper-parameters)
- Model evaluation (how well does it generalize)



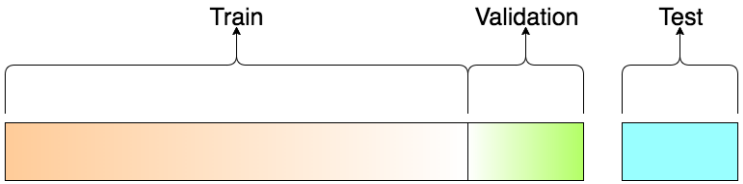
We are now potentially **overfitting** the hyper-parameters by tuning them too much. Let's look at a new approach.

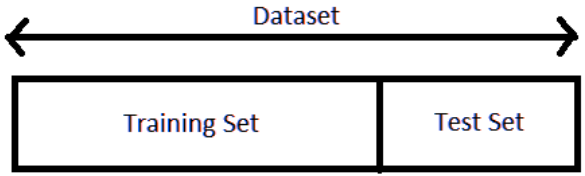
- **Training set** is used to fit the model.
- **Validation set** is used to evaluate hyper-parameter selection.
- **Test set** is used ONLY for final model evaluation.



Cross Validation

For small datasets, breaking the data up into all these groups is not idea.





Run 1	*D.tr*	*D.tr*	*D.val*
Run 2	*D.tr*	*D.val*	*D.tr*
Run 3	*D.val*	*D.tr*	*D.tr*

Idea: Divide training set into *k* groups (3 shown here) and perform *k* evaluations (where the validation set changes each time). Take the **average** of the *k* runs as the performance measure for the hyper-parameters being evaluated.

Next Steps

- Complete Lab questions 10 – 12 (15 minutes)
- Discussion

For Next time

Homework:

- Complete lab and submit to Canvas by Wed at 11:59 PM.
- Reading quiz on Canvas (due Wed at 11:59 pm)
- Work on PA 1

Reading: IDD Sections 3.4 – 3.8