# CISC 3440 Fall 2020 Final

# Part I

Due **Sunday, 12/6 11:59pm**

Share with [rebecca.levitan@share.brooklyn.edu](mailto:rebecca.levitan@share.brooklyn.edu) and denys.katerenchuk@gmail.com

## Ensemble methods

|  |  |
| --- | --- |
| What is the idea behind ensemble learning? | It uses the wisdom of the crowd; essentially, it relies on taking multiple opinions from diverse subjects to make aggregated predictions, hence improving accuracy and reducing the variance of predictions that any individual predictor would have. |
| What is hard vs. soft voting? | Hard voting is a simple poll; it counts the votes and picks the most voted. Soft voting takes into account the confidence of each voter in their answer and takes that average instead, picking the highest average of the classes. If there’s more votes for one class (“yes”) than the other, but they have low confidence in their answer, and there are almost as many votes for another (“no”), but they’re highly confident, soft voting would value that confidence and vote the latter. |
| What are 3 methods for increasing the variance of ensemble learners? Name and define each method. | 1. Bagging/Pasting - Train the predictors on random subsets of the training data, with/without replacement respectively. 2. Random Subspaces - Similar to bagging, but instead of random subsets of the instances you use random subsets of the features. 3. Random Patches - Both bagging and random subspaces; randomly sample the instances and input features for each predictor. |
| What is a Random Forest classifier? | An ensemble of Decision Tree classifiers; it trains each one on a portion of the training data and (usually) outputs the class most voted for. |
| What is Boosting? | Training predictors sequentially, with the goal of each one to improve the previous one by correcting its mistakes (hence giving relatively weak learners a boost). |
| What is Stacking? | Training (a) predictor(s) to find the best weights for the votes of the previous layer of the stack, helping us to make better use of our ensemble. |

## Dimensionality reduction

|  |  |
| --- | --- |
| What is the curse of dimensionality? | You get mathematical issues in higher dimensions that aren’t present or are insignificant in lower dimensions. For example, distances of random vectors in high dimensions are very large, which makes your model require unreasonable amounts of data not to heavily overfit these sparse data points. |
| What is an example of a dimensionality reduction technique? | PCA (principal component analysis) that reduces the dimensions as much as possible whilst keeping as much variance as asked (say, 95%). |

## Unsupervised learning

|  |  |
| --- | --- |
| What is unsupervised learning? | Learning without training labels. |
| Why does unsupervised learning have tremendous potential compared to supervised learning? | Labelling data is almost always very expensive and in much less supply than unlabeled data. If we’re able to properly utilize unlabeled data we would get a large boost in our algorithms’ success, even if unlabeled data has much lower training potential than labeled data. |
| What is the k-means clustering algorithm? | 1. Initialize k centroids to random instances 2. Label the instances (assign them to the clusters based on their distance to each centroid and picking the closest) 3. Move the centroids to the mean of all the instances of the cluster 4. Repeat b & c until the centroids stop moving |

# Part II

## Intro to deep learning

## Training DNNs

## CNNs

## GANs

## RNNs

## NLP