CSE 4820 / CSE 5819 Fall 2025-Section 001
Instructor: Jinbo Bi
Introduction to Machine Learning

Assignment 6 (100 pts)
Due: Oct 28th (Tuesday) mid-night
Support Vector Machines, Decision Trees

Name: _____ NetID: _____

Part 1: (ChatGPT Self-Learning) 20pts

This semester we are trying to use some generative AI to help learn ML concepts. Keep in mind that generative AI does not always search from authoritative websites and does not always answer the questions correctly. It is your responsibility to judge and learn from multiple sources based on our in-class discussion. You can start from the following prompts and create subsequent questions that attempt to understand the ML concepts. For each question you have asked chatGPT, just copy the first three lines of its answer into your answer sheet (so to save the space but show that you have studied). Please include this part in the end of your HW after you answer HW part 2.

1. What is the most used evaluation metrics for classification?
2. What is the most used evaluation metrics for regression?
3. For supervised learning algorithms, what would be a reliable procedure of evaluating their performance?
4. Study bagging (bootstrapping aggregation)?
5. Why the bagging procedure can create models of lower variance?
6. What are the advantages of Support vector machine?
7. What is support vector regression? (We have discussed support vector classifier, but SVM can also be extended to solve regression problems.)
8. How many different formulations of SVM?
9. How to solve an SVM optimization problem?

Part 2: Answer the Following Problems (80pts).

This part will be graded based on correctness, accuracy and clarify. Please prepare your answers in a pdf file to submit through HuskyCT assignment portal and clearly label your file with part 2 of assignment number. If you decide to use handwriting, make sure your handwriting is readable; or otherwise TAs have all rights to give 0 pt for answers that they cannot read. Please provide your calculation process, based on which you may get partial scores even if your answer is not correct.
Please try not to use ChatGPT (or other generative AI) to answer this part. If you rely too much on generative AI for this part, you may not learn enough and be well prepared for exams.

[**Support Vector Machine]** [25 pts]

(a) [10 pts] The following is the primal formulation of L2 SVM (with the squared slack variables), which is a variant of the standard SVM that we discussed in our lecture.

$$\min_{w,b,\xi} \frac{1}{2}\mathrm{w}^T\mathrm{w} + \frac{C}{2}\sum_{i=1}^{N}\xi_i^2$$
$$s.t., \quad y_i(w^T x_i + b) \geq 1 - \xi_i, \quad i \in \{1, \cdots, N\},$$
$$\xi_i \geq 0, \quad i \in \{1, \cdots, N\}.$$

If we remove the last constraints ($\xi_i \geq 0$), we might get a simpler optimization problem:

$$\min_{w,b,\xi} \frac{1}{2}\mathrm{w}^T\mathrm{w} + \frac{C}{2}\sum_{i=1}^{N}\xi_i^2$$
$$s.t., \quad y_i(w^T x_i + b) \geq 1 - \xi_i, \quad i \in \{1, \cdots, N\}.$$

Please provide the Lagrangian of the above simplified formulation.

(b) [15 pts] Please find the partial derivative of the Lagrangian in (a) with respect to w, b, and ξ_i.

[**Decision Trees**] [25 pts]

Given the dataset below you want to to build a decision tree, where x1 and x2 are the features and y is the label. This problem is about CART (please self-study and expand decision tree by checking online about CART).

x1	x2	y
Yin	7	positive
Yang	6	positive
Yin	11	positive
Yin	4	negative
Yin	1	negative
Yang	3	positive

(a) [8 pts] Which feature (x1 or x2) would you use as your first split on your decision tree, and why? (hint: think about what it means for the algorithm to "split" on a categorical feature.)

(b) [9 pts] Draw the decision tree built by the ID3-Algorithm (which is the one using entropy and we discussed in class, but you are encouraged to self-study it to further expand).

(c) [8 pts] If we were to change the labels of one of the samples in the dataset as follows:

x1	x2	y
Yin	7	positive
Yang	6	positive
Yin	11	**negative**
Yin	4	negative
Yin	1	negative
Yang	3	positive

Would the root node of the decision tree remain the same? Why?

Part 3 [**Programming**] (30 pts) In earlier HWs, you have used Seaborn Iris dataset to create an SVM classifier and you have already created codes for stratified partition of your data.

Now let us do more experiments with SVM and Decision Trees. For the following programming, please ignore virginica, meaning drop those data labeled with virginica for this problem.

(1) [15 pts] Use sklearn stratification partition to partition the Iris data into 70% in training and 30% in test. Then use the sklearn function **sklearn.svm.SVC** to create a Support Vector Classifier that separates iris species between setosa and versicolor.

Note that in HW5, we did not require you to tune the hyperparameter C (and by default, C=1), but in this HW, you will need to use a cross-validation procedure within your training data to tune the value of C. You can decide your cross-validation method (e.g., k-fold CV, 80%-20% split CV etc.). Try both linear kernel, and rbf kernel.

The CV tuning procedure is as follows:
You should have another random split within your training data or cross validation on your training data to compare the performance of SVC when choosing different values of C. Once you observe the best cross validation performance, you fix C to that optimal value, and retrain your SVC using the full training data. Then apply that final classifier to your test data. Report the classification accuracy of the final classifier on the test data and draw ROC plot with AUC computed.

(2) [15 pts] Use sklearn DecisionTreeClassiifer (https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html) with the same partition you did in (1) to create decision trees based on the 70% of the training and test on that same 30% of the testdata. Please try both "gini" and "entropy" options.

(a) For both above methods, identify the best classifiers, and then draw ROC curve using the test data, so we should see two ROC curves: one for SVC and one for decision tree.
(b) Compare the two classifiers using AUC, and examine pros and cons of SVM vs. Decision trees in terms of performance and interpretability (a small description of your observation will work.)