**CSE/ECE 343/543: Machine Learning**
**Assignment-2**

**Max Marks**: 90 (Programming:45, Theory:45) **Due Date**: 31/10/2019, 11:59PM

---

## Instructions

- Try to attempt all questions.
- Keep collaborations at high level discussions. Copying/Plagiarism will be dealt with strictly.
- Start early, solve the problems yourself. Some of these questions may be asked in Quiz/Exams.
- Late submission penalty: As per course policy.
- Your submission would be a single zip file **rollno_HW2.zip**.

---

## Programming Questions:

1. (15 points) **GridSearchCV and Support Vectors**
    1. [6 pts] Download the CIFAR 10 dataset from here. Use GridSearchCV to find the best parameters of SVM using the train set. Report the train and test accuracy and the run-times on the best parameters. Also state your observations on the best parameters. Mention the feature representation used to reduce the run-time.
    2. [6 pts] Prepare a new training set by extracting the support vectors from the previous fitted SVM (with the best parameters). Fit another SVM model with the new training set and report the train and test accuracy (Use the train data used in (i) for train accuracy).
    3. [ 3 pts] Compare the accuracy from (1) and (2). State your observations.
2. (30 points) **SVM vs Naive Bayes vs Decision Trees**
   Use the Sklearn's wine dataset and perform the following tasks. Use *train_test_split* (Scikit-learn's) to make a stratified split of 70-30 train-test with seed 42.
    1. [3 pts] Plot pairwise relations in dataset using seaborn and give inferences.
    2. [10 pts] Implement One-vs-One and One-vs-Rest classification on linear SVM. You are allowed to use only fit() and score() methods. Report all evaluation scores and your observations.
    3. [5 pts] Implement Gaussian Naive Bayes using scikit-learn and report all evaluations and any other observations that you may have.
    4. [7 pts] Use Scikit-learn's DecisionTree and tune the hyper-parameters balancing the run time and accuracy. Report all evaluations and observations.

5. [5 pts] Compare the above four models (SVM One-vs-One, SVM One-vs-Rest, Gaussian Naive Bayes and Decision Tree) on all evaluations and training time. Defend the best model for the dataset used.

Evaluations should be in the form of F-1 score, Accuracy and ROC Curve (You can use sklearn for all evaluation metrics). Additional observations may include analysis showing evidence that the model has trained correctly, class-wise accuracy, etc.

**Theory Questions**:

1. (10 points) **Convexity**

   1. [5 pts] Given a function $f : \mathbb{R}^n \to R$ is continuously differentiable. Also, the $f(x)$ is concave on a convex set $\xi$. Given the above propoerties of $f(x)$, prove that for all $x_1, x_2 \in \xi$, $f(x)$ satisfies the following property:

   $$f(x_2) \leq f(x_1) + \nabla f(x_1)(x_2 - x_1) \tag{1}$$

   here, $\nabla$ is the differential operator $\frac{d}{dx}$.
   **Hint:** Use basic definition of a derivative. Note that x is a vector, and $\nabla f(x_1)$ would be the gradient of $f$ computed at $x_1$.

   2. [5 pts] Find the set of values of $\alpha$ such that the given function is convex.

   $$f(x, y, z) = x^2 + y^2 + 5z^2 - 2xz + 2\alpha xy + 4yz \tag{2}$$

2. (35 points) **SVM**

   Suppose you are given given a data set of *six* one-dimensional data points: Three of the six data points have negative label and the other three have positive label:

   - **Negative Labels:** $x_1 = -1$, $x_2 = 0$, $x_3 = 1$
   - **Positive Labels:** $x_4 = -3$, $x_5 = -2$, $x_6 = 3$

   Plot the dataset. You may find that the data is not linearly separable. However, if we apply the feature map $\phi(u) = (u, u^2)$, the points in $\mathbb{R}^1$ will be transformed to new points in $\mathbb{R}^2$. Visualize the transformed data points in $\mathbb{R}^2$, now a linear separator can separate the points in $\mathbb{R}^2$ .

   1. [2 pts] Give the analytic form of the kernel that corresponds to the feature map $\phi$ in terms of only $X_1$ and $X_1'$. Specifically define $k(X_1, X_1')$.

   2. [10 pts] Construct a maximum-margin separating hyperplane. This hyperplane will be a line in $\mathbb{R}^2$, which can be parameterized by its normal equation, i.e. $w_1 Y_1 + w_2 Y_2 + c = 0$ for appropriate choices of $w_1, w_2, c$. Here, $(Y_1, Y_2) = \phi(X_1)$ is the result of applying the feature map $\phi$ to the original feature $X_1$. Give the values for $w_1, w_2, c$. Also, explicitly compute the margin for your hyperplane. Note that the line must pass somewhere between (-2,4) and (-1,1), and that the hyperplane must be perpendicular to the line connecting these two points Use only two support vectors.

3. [4 pts] Apply $\phi$ to the data and plot the points in the new $\mathbb{R}^2$ feature space. On the plot of the transformed points, plot the separating hyperplane and the margin, and circle the support vectors.
4. [2 pts] Draw the decision boundary of the separating hyperplane in the original $\mathbb{R}^1$ feature space.
5. [10 pts] Compute the coefficients $\alpha$ and the constant $b$ in Eq. (3) for the kernel $k$ and the support vectors $SV = \{u_1, u_2\}$ you chose in part 4. Be sure to explain how you obtained these coefficients.

$$y(x) = \text{sign}\left(\sum_{n=1}^{|SV|} \alpha_n y_n k(x, u_n) + b\right) \tag{3}$$

   Think about the dual form of the quadratic program and the constraints placed on the $\alpha$ values.
6. [2 pts] If we add another positive ($Y = +$) point to the training set at $X_1 = 5$ would the hyperplane or margin change? Why or why not?