ISE - 364/464: Introduction to Machine Learning
Homework Assignment 6

The goal of this assignment is to provide a series of problems that strengthen knowledge in support vector machines, hyperplanes, margins, kernel functions, as well as identifying certain ML model equivalences.

Grading: This assignment is due on Coursesite by E.O.D. 11/29/2024. All problems are worth the same number of points. If a problem has multiple parts, each of those parts will be worth equal amounts and will sum to the total number of points of the original problem (Example: If each problem is worth a single point, and problem 1 has 4 parts, each part will be worth 1/4th of a point). ISE - 364 students are only required to answer problems 1 through 4; however, you are allowed to answer the 5th graduate-level question (if done so correctly, you will receive extra credit in the amount that the 5th problem will be worth for the ISE - 464 students). ISE - 464 students are required to answer all 5 problems.

Submitting: Only electronic submissions on Coursesite are accepted.

1 Problems

1. (**Dual Objective Function for Hard-SVM**) Given a dataset $\mathcal{D} = \left\{\left(x^{(i)}, y^{(i)}\right)\right\}_{i=1}^{m}$ where $x^{(i)} \in \mathbb{R}^n$ and $y^{(i)} \in \{1, -1\}$, recall that the primal optimization problem for the Hard-Margin formulation of SVM is given as

$$\begin{aligned} \min_{w \in \mathbb{R}^n,\, b \in \mathbb{R}} \quad & \frac{1}{2}\|w\|^2 \\ \text{s.t.} \quad & y^{(i)} \left(w^\top x^{(i)} - b\right) \geq 1, \qquad \forall i \in \{1, 2, ..., m\}. \end{aligned}$$

   Prove that the dual objective function (the objective function for the dual formulation of Hard-SVM) is

$$\hat{\mathcal{L}}(\alpha) := \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} \left({x^{(i)}}^\top x^{(j)}\right),$$

   where the Lagrange multipliers satisfy $\alpha_i \geq 0$ for all $i \in \{1, 2, ..., m\}$.
   (Hint: Refer to the SVM lecture slides.)

2. (**Deriving the Dual Problem for Soft-SVM**) Given a dataset $\mathcal{D} = \left\{\left(x^{(i)}, y^{(i)}\right)\right\}_{i=1}^{m}$ where $x^{(i)} \in \mathbb{R}^n$ and $y^{(i)} \in \{1, -1\}$, recall that the primal optimization problem for the

Soft-Margin formulation of SVM, given some constant $C \geq 0$, is defined as

$$
\begin{aligned}
\min_{w \in \mathbb{R}^n,\, b \in \mathbb{R}} \quad & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} \xi_i \\
\text{s.t.} \quad & y^{(i)}\left(w^\top x^{(i)} - b\right) \geq 1 - \xi_i, && \forall i \in \{1, 2, ..., m\}, \\
& \xi_i \geq 0, && \forall i \in \{1, 2, ..., m\}.
\end{aligned}
$$

Prove that the corresponding dual problem formulation of Soft-SVM is

$$
\begin{aligned}
\max_{\alpha \in \mathbb{R}^m,\, \mu \in \mathbb{R}^m} \quad & \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} \left({x^{(i)}}^\top x^{(j)}\right) \\
\text{s.t.} \quad & \sum_{i=1}^{m} \alpha_i y^{(i)} = 0, \\
& C - \alpha_i - \mu_i = 0, && \forall i \in \{1, 2, ..., m\}, \\
& \mu_i \geq 0, && \forall i \in \{1, 2, ..., m\}, \\
& \alpha_i \geq 0, && \forall i \in \{1, 2, ..., m\}.
\end{aligned}
$$

(Hint: Refer to the SVM lecture slides.)

3. (**Hyperplanes, Functional Margins, and Geometric Margins**) Consider the three training examples $\left(x^{(1)}, y^{(1)}\right) = \left([3,4]^\top, 1\right)$, $\left(x^{(2)}, y^{(2)}\right) = \left([1,-1]^\top, -1\right)$, and $\left(x^{(3)}, y^{(3)}\right) = \left([-2,-3]^\top, -1\right)$. Further, consider the hyperplane given by $h(x; w, b) = w^\top x - b$, where $w = [1,2]^\top$ and $b = 1$.

- **a)** Compute the functional margins and geometric margins of these three datapoints with respect the the hyperplane h.
- **b)** Compute the normalized normal vector $\frac{w}{\|w\|}$ as well as the distance from the origin to the hyperplane.
- **c)** Plot the three datapoints in $\mathbb{R}^2$ (colored by their target class) as well as the hyperplane h.
- **d)** Does the hyperplane correctly classify the three datapoints? Is this hyperplane the "best-fit" (i.e., does this hyperplane yield the largest margins among the support vectors)?

4. (**Evaluating Kernel Functions**) Consider the two training examples $x^{(1)} = [3,4]^\top$ and $x^{(2)} = [1,-1]^\top$.

- **a)** Compute the value of the Linear Kernel function on these two datapoints.
- **b)** Compute the value of the Gaussian Kernel function, given that $\sigma = 1.5$, on these two datapoints.

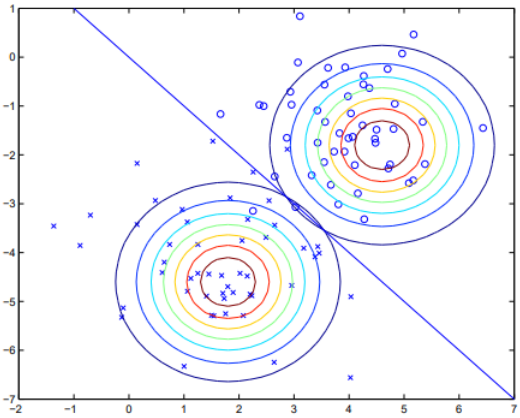

Figure 1: Illustration of a GDA model for two multivariate Gaussian distributions with different means but equal covariances.

5. (ISE-464 Graduate Students) (**Logistic Regression and GDA**) Gaussian Discriminant Analysis (GDA) is a generative machine learning model that utilizes Baye's rule to calculate posterior distribution probabilities

$$\mathbb{P}(y|x) = \frac{\mathbb{P}(y)\mathbb{P}(x|y)}{\mathbb{P}(x)},$$

where $\mathbb{P}(x|y)$ is a multivariate Gaussian distribution given by

$$\mathbb{P}(x|y=c) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu_c)^\top \Sigma^{-1}(x-\mu_c)},$$

where Σ is the covariance matrix and μ_c is the mean of Gaussian corresponding to the c-th target class. Further, assume that $x \in \mathbb{R}^n$ and $y \in \{0,1\}$ (i.e., the target variable is binary). This means that $\mathbb{P}(x|y=0)$ and $\mathbb{P}(x|y=1)$ are multivariate Gaussian distributions with means of μ_0 and μ_1 and covariance of Σ (for both), respectively. The GDA model for this type of data can be visualized in Figure 1.

Prove that a GDA model with two target classes and the same covariance Σ for both classes' Gaussians, leads to a logistic regression model. That is, prove that $\mathbb{P}(y=1|x) = \sigma(z)$, where σ is the sigmoid function and z is some linear function.

(Hint: Notice that one can write $\mathbb{P}(y=1|x) = \frac{\mathbb{P}(y=1)\mathbb{P}(x|y=1)}{\mathbb{P}(x)} = \frac{1}{1+\exp\left\{\log \frac{\mathbb{P}(y=0)\mathbb{P}(x|y=0)}{\mathbb{P}(y=1)\mathbb{P}(x|y=1)}\right\}}$*.)*