Part 1. Theory

A. Linear Classifier Margin (20 POINTS)

Given n orthogonal vectors of the form $(0, ..., 0, 1, 0, ..., 0)$, show that for any labelling $\{+1, -1\}$ of the vectors, there always exists a linear classification rule with margin at least $\frac{1}{\sqrt{n}}$.

Consider $\vec{w} = \sum_{j=1}^{n} y_j * \vec{x}_j$ and $b = 0$. This hyperplane classifies correctly as:

$$\begin{array}{l} \forall i \;\; y_i * (\vec{w} * \vec{x}_i + b) = \\ y_i * \sum_{j=1}^{n} y_j * \vec{x}_j * \vec{x}_i = \\ y_i * y_i = 1 \end{array}$$

Also $\vec{w}$ is a vector with at most n ones in it and therefore $||\vec{w}|| <= \sqrt{n}$ and so the margin for $\vec{w}$ is at least $\frac{1}{\sqrt{n}}$.

B. SVM Hard-Margin Classification (20 POINTS)

Show that one can replace the "1" on the right hand side of the constraint in hard-margin SVMs (repeated below) with any positive number $\gamma > 0$ without changing the hyperplane that is computed.

$$\begin{array}{l} \min_{\vec{w}, b} \frac{1}{2} ||\vec{w}||^2 \\ \text{subject to } y_i(\vec{x}_i \cdot \vec{w} + b) \geq 1 \; \forall i \in 1..n, \end{array}$$

Let $\vec{w}_1, b_1$ and $\vec{w}_\gamma, b_\gamma$ be the solutions to the probelms with 1 and γ respectively on the right hand side. If we substitute $\frac{\vec{w'}}{\gamma}, \frac{b'}{\gamma}$ for $\vec{w}, b$ we get:

$$\begin{array}{l} \min_{\frac{\vec{w'}}{\gamma}, \frac{b'}{\gamma}} \frac{1}{2} ||\vec{w'}||^2 \\ \text{subject to } \; y_i(\vec{x}_i \cdot \vec{w'} + b') \geq \gamma \forall i \in 1..n, \end{array}$$

as the $\frac{1}{\gamma^2}$ in the minimization is just a positive constant. But this is the minimization that we get when we have γ on the right hand side instead of 1. So we get that $\frac{\vec{w}_1}{\gamma} = \vec{w}_\gamma$ and $\frac{b_1}{\gamma} = b_\gamma$. Now these two values of $\vec{w}$ and b define the same plane as

$$\begin{array}{l} \vec{w}_1 * \vec{x} + b_1 = 0 \iff \\ \frac{\vec{w}_1}{\gamma} * \vec{x} + \frac{b_1}{\gamma} = 0 \iff \\ \vec{w}_\gamma * \vec{x} + b_\gamma = 0 \end{array}$$

C. SVM Soft-Margin Classification (20 POINTS)

Imagine you want to train the following soft-margin SVM, where the slack variables enter the objective function squared.

$$\begin{array}{l} \min_{\vec{w},\, b,\, \vec{\xi}} \frac{1}{2} ||\vec{w}||^2 + C \sum_{i=1}^{n} \xi_i^2 \\ \text{subject to } y_i(\vec{x}_i \cdot \vec{w} + b) \geq 1 - xi_i \; \forall i \in 1..n \end{array}$$

However, you do not have quadratic programming software that can solve this optimization problem. All you have is a program that can solve hard-margin problems of the form.

$$\begin{array}{l} \min_{\vec{w},\, b} \frac{1}{2} ||\vec{w}||^2 \\ \text{subject to } y_i(\vec{x}_i \cdot \vec{w} + b) \geq 1 \; \forall i \in 1..n, \end{array}$$

Show how you can transform the soft-margin problem into an equivalent problem that can be trained with the hard-margin software. Hint: Extend the feature vectors $\vec{x}_i$ by adding features in the appropriate way.

Let $\vec{x'}_i = (\vec{x}_i \;\; \frac{y_i}{\sqrt{2C}} * \vec{e}_i)$ where $\vec{e}_i$ is i^{th} row of the $n \times n$ identity matrix. Let $\vec{w'}$ and b' be the results of solving the hard-margin problem with these vectors. Now if we set $\vec{w}_i = \vec{w'}_i \;\; \forall i \in 1..n$ and $\xi_i = \frac{1}{\sqrt{2C}} * \vec{w'}_{n+i} \;\; \forall i \in 1..n$ then we can see that from the solution to the hard margin problem we have the gaurantee that $\vec{w'}$ minimizes $\frac{1}{2}||\vec{w}[1..2n]||^2$ and this is equivalent to saying that $\vec{w}, \vec{\xi}$ minimizes $\frac{1}{2}||\vec{w}[1..n]||^2 + C \sum_{i=1}^{n} \xi_i^2$ subject to

$$
\begin{aligned}
& y_i(\vec{x'}_i \cdot \vec{w'} + b) \geq 1 \;\; \forall i \in 1..n \iff \\
& y_i((\vec{x}_i \;\; \tfrac{y_i}{\sqrt{2C}} * \vec{e}_i) \cdot (\vec{w} \;\; \sqrt{2C} * \vec{\xi}) + b) \geq 1 \;\; \forall i \in 1..n \iff \\
& y_i(\vec{x}_i \cdot \vec{w}_i + y_i * \xi_i + b) \geq 1 \;\; \forall i \in 1..n \iff \\
& y_i(\vec{x}_i \cdot \vec{w}_i + b) \geq 1 - \xi_i \;\; \forall i \in 1..n
\end{aligned}
$$

which is exactly what we want.