Math 342W/642/742W

Recitation – Day #4 (2.11.25)

I. Hyperplanes

(i) Describe/explain/define what a **hyperplane** is in the context of $\mathbb{R}^2$ or $\mathbb{R}^3$.
(*Visualizations may be useful.*)

A **hyperplane** is a 1-dimensional plane (a line) in $\mathbb{R}^2$ and its a 2-dimensional plane in $\mathbb{R}^2$. The commonality of these hyperplanes in either context is that the line or plane divides/partitions $\mathbb{R}^2$ or $\mathbb{R}^3$ into two (*half*-)spaces, respectively.

(ii) Describe/explain/define what a **hyperplane** is in the context of $\mathbb{R}^n$.

In $\mathbb{R}^n$, a **hyperplane** is a $(n-1)-$dimensional generalization of a 2D-plane in $\mathbb{R}^3$.

(iii) Write the "**Hesse Normal Form**" definition of a *hyperplane*. What does it mean that it is *overparametrized*?

Hesse Normal Form of a Hyperplane: $\boldsymbol{w} \cdot \boldsymbol{x} - b = 0$

It is considered to be *overparametrized* because for any nonzero constant c,

$$c\,\boldsymbol{w} \cdot \boldsymbol{x} - cb = 0$$

represents the same hyperplane as $\boldsymbol{w} \cdot \boldsymbol{x} - b = 0$.

II. Terminology of Hyperplanes

(i) Define the set of candidate functions of interest, $\mathcal{H}$.

$$\mathcal{H} = \{\mathbb{1}_{\boldsymbol{w}\cdot\boldsymbol{x}-b\geq 0} \,:\, \boldsymbol{w} \in \mathbb{R}^p,\, b \in \mathbb{R}\}$$

(ii) Define the following terms:

- $\boldsymbol{w}$: column vector of p weights ($\boldsymbol{w} \in \mathbb{R}^p$) which is a normal vector perpendicular to $\boldsymbol{\ell}$
- $\boldsymbol{w}_0$: $= \dfrac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$ (unit vector of $\boldsymbol{w}$)
- b: bias term
- $\boldsymbol{\ell}$: the *optimal* hyperplane
- $\boldsymbol{\ell}_L$: the lower boundary of the wedge/margin parallel to $\boldsymbol{\ell}$
- $\boldsymbol{\ell}_U$: the upper boundary of the wedge/margin parallel to $\boldsymbol{\ell}$
- $\boldsymbol{z}$: $\dfrac{b}{\|\boldsymbol{w}\|}\boldsymbol{w}_0$ (vector on $\boldsymbol{\ell}$)
- $\boldsymbol{z}_L$: $\dfrac{b-\delta}{\|\boldsymbol{w}\|}\boldsymbol{w}_0$ (vector on $\boldsymbol{\ell}_L$)
- $\boldsymbol{z}_U$: $\dfrac{b+\delta}{\|\boldsymbol{w}\|}\boldsymbol{w}_0$ (vector on $\boldsymbol{\ell}_U$)
- $m := \dfrac{2}{\|\boldsymbol{w}\|}$ (width of margin)

III. Finding the Optimal Hyperplane

(i) What is the main goal in finding the optimal hyperplane among linearly separable data?

Finding the optimal hyperplane among linearly separable data points can be visualized in $\mathbb{R}^2$ where the line that divides the largest/widest "wedge" between two linearly separable sets of data points is considered as optimal. We then extend this geometric notion of a optimal hyperplane in $\mathbb{R}^2$ to $\mathbb{R}^n$.

(ii) Which quantity are we *maximizing*? Which quantity are we *minimizing*?

To find the optimal hyperplane we have to:

- *maximize* m (width of the margin/wedge between the two groups of data)
- *minimize* $\|\boldsymbol{w}\|$ (the norm of the vector of weights)

(iii) How do we express the optimization problem for finding the optimal hyperplane among linearly separable data?

Optimization Problem for finding the optimal hyperplane:

$$\text{Minimize } \|\boldsymbol{w}\| \text{ such that } \left(y_i - \tfrac{1}{2}\right) \cdot (\boldsymbol{w} \cdot \boldsymbol{x} - b) \geq \tfrac{1}{2}, \quad \forall i \in \{1, \dots, n\}$$

IV. Support Vector Machines (SVM)

(i) What constitutes as an **error** if we want to find the optimal hyperplane for when data are not linearly separable?

Errors arise when the two sets of data we are trying to divide by a line are ***not*** linearly separable so that at least one data point that will be misclassified by any proposed hyperplane.

(ii) Define **Hinge Error** (HE)/**total Hinger Error**(THE).

Hinge Error (HE):

$$\text{HE}_i = \max\left\{0, \frac{1}{2} - \left(y_i - \frac{1}{2}\right) \cdot (\boldsymbol{w} \cdot \boldsymbol{x}_i - b)\right\}$$

Total Hinge Error (THE):

$$\text{THE} = \sum_{i=1}^{n} \max\left\{0, \frac{1}{2} - \left(y_i - \frac{1}{2}\right) \cdot (\boldsymbol{w} \cdot \boldsymbol{x}_i - b)\right\}$$

(iii) Express/formulate the optimization problem for finding the optimal hyperplane using *Vapnik* objective/fitness/loss function to be minimized.

$$\underset{\boldsymbol{w}, b}{\text{argmin}} \left\{\frac{1}{2} \sum_{i=1}^{n} \max\left\{0, \frac{1}{2} - \left(y_i - \frac{1}{2}\right) \cdot (\boldsymbol{w} \cdot \boldsymbol{x}_i - b)\right\} + \lambda \|\boldsymbol{w}\|^2\right\} \quad (\text{hyperparmater: } \lambda > 0)$$

(iv) What category of problems are SVMs applicable for?

SVMs are applicable/suitable for ***binary classification*** problems.