# CM146, Fall 2018
# Problem Set 2: Perceptron and Regression

Rodrigo Valle

31 October 2018

## 1 Problem 1: Perceptron

**Solution:** Please check CCLE to see answer for this question.

## 2 Problem 2: Logistic Regression

**Solution:** Please check CCLE to see answer for this question.

## 3 Problem 3: Understanding Linear Separability

**Proposed Linear Program:**

$$\begin{aligned} \min \quad & \delta \\ \text{subject to} \quad & y_i(\mathbf{w}^T\mathbf{x}_i + \theta) \geq 1 - \delta \quad \forall (\mathbf{x}_i, y_i) \in D \\ & \delta \geq 0 \end{aligned}$$

(a) A data set $D = \{(\mathbf{x_i}, y_i)\}_{i=1}^m$ that satisfies condition (1) above is called *linearly separable*. Show that there is an optimal solution with $\delta = 0$, then $D$ is linearly separable.

**Solution:**

i. **Solution to Linear Program with $\delta = 0 \Rightarrow$ Linear Separability**

If we plug $\delta = 0$ into our constraints, we have that our hyperplane solution to the linear program satisfies:

$$y_i(\mathbf{w}^T\mathbf{x}_i + \theta) \geq 1$$

If $D$ is linearly separable, then $\mathbf{w}$ and $\theta$ are constrained such that:

$$y_i = \begin{cases} 1 & \text{if } \mathbf{w}^T\mathbf{x}_i + \theta \geq 0 \\ -1 & \text{if } \mathbf{w}^T\mathbf{x}_i + \theta < 0 \end{cases} \forall (\mathbf{x_i}, y_i) \in D$$

We observe that this is equivalent to stating:

$$y_i(\mathbf{w}^T\mathbf{x}_i + \theta) \geq 0 \quad \forall (\mathbf{x}_i, y_i) \in D$$

So given these constraints, with $\delta = 0$ we will always predict a value greater than or equal to one, which is always greater than or equal to zero, which indicates that our hyperplane solution with $\delta = 0$ separates $D$.

$$y_i(\mathbf{w}^T\mathbf{x}_i + \theta) \geq 1 \geq 0 \quad \forall (\mathbf{x}_i, y_i) \in D$$

ii. **Linear Separability $\Rightarrow$ Solution to Linear Program with $\delta = 0$**

For completeness, as the question seems unclear with conflicting information from Piazza, we'll also show that linear separability implies $\delta = 0$.

If $D$ is separable, then there exist infinitely many separating hyperplanes. Let's select one of these separating hyperplanes: $\mathbf{a}^T\mathbf{x} + b$, so we have that:

$$y_i(\mathbf{a}^T\mathbf{x}_i + b) \geq 0 \quad \forall (\mathbf{x}_i, y_i) \in D$$

To solve the linear program, we want to choose $\mathbf{a}$ and $b$ so that $\delta = 0$. That is, we want our hyperplane to separate every example in $D$ with a margin (after projection) greater than or equal to 1.

$$y_i(\mathbf{a}^T\mathbf{x}_i + b) \geq 1 \quad \forall(\mathbf{x}_i, y_i) \in D$$

We find distance $d$ (after projection) to the closest point to our hyperplane:

$$d = \min_{(\mathbf{x}_i, y_i)\in D} y_i(\mathbf{a}^T\mathbf{x}_i + b)$$

so that we can be sure that

$$y_i(\mathbf{a}^T\mathbf{x}_i + b) \geq d \quad \forall(\mathbf{x}_i, y_i) \in D$$

and we divide through to normalize the size of our projection

$$\frac{y_i(\mathbf{a}^T\mathbf{x}_i + b)}{d} \geq 1 \quad \forall(\mathbf{x}_i, y_i) \in D$$

and get our new $\mathbf{w}$ and $\theta$:

$$\mathbf{w} = \frac{\mathbf{a}}{d}, \quad \theta = \frac{b}{d}$$

Now we have a separating hyperplane that solves our linear program subject to the constraint $\delta = 0$ such that:

$$y_i(\mathbf{w}^T\mathbf{x}_i + \theta) \geq 1 \quad \forall(\mathbf{x}_i, y_i) \in D$$

In the case that $d = 0$, i.e. an example with $y_i = 1$ lies directly on the separating hyperplane, we should adjust our separating hyperplane so that no examples lie on it. We can find such a hyperplane by moving our original choice of hyperplane to the center of the closest positive and negative examples

$$\begin{aligned} d^+ &= \min_{(\mathbf{x}_i, y_i)\in\{D|y_i=1\}} (\mathbf{a}^T\mathbf{x}_i + b) \\ d^- &= \max_{(\mathbf{x}_i, y_i)\in\{D|y_i=-1\}} (\mathbf{a}^T\mathbf{x}_i + b) \end{aligned}$$

Our new hyperplane will be $\mathbf{a}^T\mathbf{x}_i + b - \frac{d^+ + d^-}{2} = 0$, which still separates $D$ because we have translated it along the opposite direction of $\mathbf{a}$ (towards the negative examples) by a distance which is not far enough to cause us to misclassify a negative example. That is:

$$d^+ - \frac{d^+ + d^-}{2} \geq 0 > d^- - \frac{d^+ + d^-}{2}$$

Our adjusted hyperplane then is subject to the following constraint:

$$y_i\left(\mathbf{a}^T\mathbf{x}_i + b - \frac{d^+ + d^-}{2}\right) \geq \frac{d^+ + d^-}{2} \quad \forall(\mathbf{x}_i, y_i) \in D$$

normalizing as before gives us

$$\mathbf{w} = \frac{\mathbf{a}}{\frac{d^+ + d^-}{2}}, \quad \theta = \frac{b}{\frac{d^+ + d^-}{2}} - 1$$

which gives a solution to the linear program with minimal $\delta = 0$ as before.

$$y_i(\mathbf{w}^T\mathbf{x}_i + \theta) \geq 1 \quad \forall(\mathbf{x}_i, y_i) \in D$$

(b) What can we say about the linear separability of the data set if there exists a hyperplane that satisfies condition (2) with $\delta > 0$?

**Solution:** The data is not necessarily separable. If it's also true that $\delta < 1$, then we can use a process like the one in part (a) to show that $D$ is linearly separable.

If $\delta$ is minimal and $\delta \geq 1$, then we can be certain that $D$ is not linearly separable. On the other hand, if $\delta \geq 1$ but $\delta$ is not minimal, then we cannot conclude if $D$ is linearly separable or not because there could exist another hyperplane with $\delta < 1$.

(c) An alternative LP formulation to (2) may be

$$\begin{aligned} \min \quad & \delta \\ \text{subject to} \quad & y_i(\mathbf{w}^T\mathbf{x}_i + \theta) \geq -\delta \quad \forall(\mathbf{x}_i, y_i) \in D \\ & \delta \geq 0 \end{aligned}$$

Find the optimal solution to this formulation (independent of $D$) to illustrate the issue with such a formulation.

**Solution:** A possible solution that would work for any $D$ is $\mathbf{w} = \vec{0}$, $\theta = 0$, $\delta = 0$. $\delta$ is minimized, and all constraints are satisfied. This formulation, however, will not help us find a separating hyperplane because it merely collapses every example into the origin upon projection to satisfy its constraints — note that it will work even for $D$ that are not linearly separable.

(d) Let $\mathbf{x}_1 \in \mathbb{R}^n, \mathbf{x}_1^T = [1\ 1\ 1]$ and $y_1 = 1$. Let $\mathbf{x}_2 \in \mathbb{R}^n, \mathbf{x}_2^T = [-1\ \ -1\ \ -1]$ and $y_2 = -1$. The data set $D$ is defined as

$$D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2)\}$$

Consider the formulation in (2) applied to $D$. What are possible optimal solutions?

**Solution:** $|D| = 2$ and the two examples are not at the same point, so $D$ is trivially linearly separable — meaning there exist optimal solutions to the linear program with $\delta = 0$.

So, we want to choose $\mathbf{w}$ and $\theta$ to satisfy the constraint

$$y_i(\mathbf{w}^T \mathbf{x}_i + \theta) \geq 1 \quad \forall (\mathbf{x}_i, y_i) \in D$$

Expanded, we have the constraints

$$\begin{aligned} w_1 + w_2 + w_3 + \theta \geq 1 &\quad \text{for } \mathbf{x}_1 \\ w_1 + w_2 + w_3 - \theta \geq 1 &\quad \text{for } \mathbf{x}_2 \end{aligned}$$

which we can simplify to the dominating constraint

$$w_1 + w_2 + w_3 \geq 1 + |\theta|$$

which defines a space where all optimal solutions live.

We can select an optimal solution from this space, such as

$$\mathbf{w} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \theta = 0$$

which separates our data nicely.