Compsci 571 HW3

Yilin Gao (yg95)

February 9, 2018

1 Separability

Proof:

To show that two convex hulls do not intersect, we could show any point in one convex hull is not in the other convex hull, or equivalently, any point in one convex hull is not the same as any point in the other convext hull.

Suppose $\mathbf{x}_i = \sum_n \alpha_{ni}\mathbf{x}_n$ is a point in the convex hull of $\{\mathbf{x}_n\}$, with $\alpha_{ni} \geq 0$ for $\forall i$ and $\sum_n \alpha_{ni} = 1$, and $\mathbf{x}_j = \sum_m \alpha_{mj}\mathbf{x}'_m$ is a point in the convex hull of $\{\mathbf{x}'_m\}$, with $\alpha_{mj} \geq 0$ for $\forall j$ and $\sum_m \alpha_{mj} = 1$. We need to prove $\mathbf{x}_i \neq \mathbf{x}'_j$ for $\forall i$ and $\forall j$.

We prove with contradiction. Assume for a given i, there exists some j such that $\mathbf{x}_i = \mathbf{x}'_j$. Then $\mathbf{w}^T\mathbf{x}_i = \mathbf{w}^T\mathbf{x}'_j$, and $\mathbf{w}^T\mathbf{x}_i - \mathbf{w}^T\mathbf{x}'_j = 0$.

$$\mathbf{w}^T\mathbf{x}_i - \mathbf{w}^T\mathbf{x}'_j = \mathbf{w}^T(\sum_n \alpha_{ni}\mathbf{x}_n) - \mathbf{w}^T(\sum_m \alpha_{mj}\mathbf{x}'_m) = \sum_n \alpha_{ni}(\mathbf{w}^T\mathbf{x}_n) - \sum_m \alpha_{mj}(\mathbf{w}^T\mathbf{x}'_m)$$

$$= [\sum_n \alpha_{ni}(\mathbf{w}^T\mathbf{x}_n) + w_0] - [\sum_m \alpha_{mj}(\mathbf{w}^T\mathbf{x}'_m) + w_0]$$

Because $\sum_n \alpha_{ni} = 1$ and $\sum_m \alpha_{mj} = 1$:

$$= [\sum_n \alpha_{ni}(\mathbf{w}^T\mathbf{x}_n) + \sum_n \alpha_{ni}w_0] - [\sum_m \alpha_{mj}(\mathbf{w}^T\mathbf{x}'_m) + \sum_m \alpha_{mj}w_0]$$

$$= \sum_n \alpha_{ni}(\mathbf{w}^T\mathbf{x}_n + w_0) - \sum_m \alpha_{mj}(\mathbf{w}^T\mathbf{x}'_m + w_0)$$

If the two sets of points $\{\mathbf{x}_i\}$ and $\{\mathbf{x}'_j\}$ are linearly seprable, $\mathbf{w}^T\mathbf{x}_n + w_0 > 0$ for $\forall n$, and $\mathbf{w}^T\mathbf{x}'_m + w_0 < 0$ for $\forall m$. And because $\alpha_{ni} \geq 0$ for $\forall ni$ and $\alpha_{mj} \geq 0$ for $\forall mj$, the above equation is a sum of n positive numbers minus a sum of m negative numbers, which is a positive number.

So $\mathbf{w}^T\mathbf{x}_i$ cannot be equal to $\mathbf{w}^T\mathbf{x}'_j$. Contradiction.

So for a given i, there is no j such that $\mathbf{x}_i = \mathbf{x}'_j$. So the two convex hulls do not intersect. $\square$

2 Logistic Regression and Gradient Descent

(a) *Proof:*

$$\sigma'(a) = -\frac{1}{(1+e^{-a})^2}(-e^{-a}) = \frac{e^{-a}}{(1+e^{-a})^2}$$

$$\sigma(a)(1-\sigma(a)) = \frac{1}{1+e^{-a}}(1 - \frac{1}{1+e^{-a}}) = \frac{1}{1+e^{-a}}\frac{e^{-a}}{1+e^{-a}} = \frac{e^{-a}}{(1+e^{-a})^2}$$

$$\therefore \sigma'(a) = \sigma(a)(1-\sigma(a)) \quad \square$$

(b)

$$\frac{\partial L_{\mathbf{w}}}{\partial w_j} = \frac{\partial}{\partial w_j} \sum_{i=1}^{n} \{-y^{(i)} \log[\sigma(\mathbf{w}^T \mathbf{x}^{(i)})] - (1 - y^{(i)}) \log[1 - \sigma(\mathbf{w}^T \mathbf{x}^{(i)})]\}$$

$$= \sum_{i=1}^{n} \{-y^{(i)} \frac{\sigma'(\mathbf{w}^T \mathbf{x}^{(i)})}{\sigma(\mathbf{w}^T \mathbf{x}^{(i)})} x_j^{(i)} - (1 - y^{(i)}) \frac{-\sigma'(\mathbf{w}^T \mathbf{x}^{(i)})}{1 - \sigma(\mathbf{w}^T \mathbf{x}^{(i)})} x_j^{(i)}\}$$

Because $\sigma'(\mathbf{w}^T \mathbf{x}) = \sigma(\mathbf{w}^T \mathbf{x})(1 - \sigma(\mathbf{w}^T \mathbf{x}))$,

$$= \sum_{i=1}^{n} \{-y^{(i)}(1 - \sigma(\mathbf{w}^T \mathbf{x}^{(i)})) x_j^{(i)} + (1 - y^{(i)}) \sigma(\mathbf{w}^T \mathbf{x}^{(i)}) x_j^{(i)}\}$$

$$= \sum_{i=1}^{n} \{x_j^{(i)}(\sigma(\mathbf{w}^T \mathbf{x}^{(i)}) - y^{(i)})\}$$

(c) To prove $L_{\mathbf{w}}$ is convex in $\mathbf{w}$, we need to prove that for any $\mathbf{w}_1$ and $\mathbf{w}_2$ and $\theta \in [0,1]$, $L(\theta \mathbf{w}_1 + (1-\theta)\mathbf{w}_2) \leq \theta L(\mathbf{w}_1) + (1 - \theta)L(\mathbf{w}_2)$.

Use $\mathbf{w}^*$ to represent $\theta \mathbf{w}_1 + (1 - \theta)\mathbf{w}_2$.

For the left-hand side:

$$LHS = L(\mathbf{w}^*) = \sum_{i=1}^{n} \{-y^{(i)} \log[\sigma(\mathbf{w}^{*T} \mathbf{x})] - (1 - y^{(i)}) \log[1 - \sigma(\mathbf{w}^{*T} \mathbf{x})]\}$$

Because $\sigma(\mathbf{w}^{*T} \mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}^{*T} \mathbf{x}}} = \frac{1}{1 + e^{-\theta \mathbf{w}_1^T \mathbf{x}} * e^{-(1-\theta)\mathbf{w}_2^T \mathbf{x}}}$, define $A \triangleq 1 + e^{-\theta \mathbf{w}_1^T \mathbf{x}} * e^{-(1-\theta)\mathbf{w}_2^T \mathbf{x}}$,

$$= \sum_{i=1}^{n} \{y^{(i)} \log A - (1 - y^{(i)})(\log(1 - A) - \log A)\}$$

$$= \sum_{i=1}^{n} \{y^{(i)} \log A + (1 - y^{(i)})(\mathbf{w}^{*T} \mathbf{x} + \log A)\}$$

$$= \sum_{i=1}^{n} \{(1 - y^{(i)})\mathbf{w}^{*T} \mathbf{x} + \log A)\}$$

For the right hand side:

$$RHS = \theta L(\mathbf{w}_1) + (1 - \theta)L(\mathbf{w}_2)$$

$$= \theta \sum_{i=1}^{n} \{-y^{(i)} \log[\sigma(\mathbf{w}_1^T \mathbf{x})] - (1-y^{(i)}) \log[1-\sigma(\mathbf{w}_1^T \mathbf{x})]\} + (1-\theta) \sum_{i=1}^{n} \{-y^{(i)} \log[\sigma(\mathbf{w}_2^T \mathbf{x})] - (1-y^{(i)}) \log[1-\sigma(\mathbf{w}_2^T \mathbf{x})]\}$$

$$= \sum_{i=1}^{n} \{-y^{(i)}[\theta \log[\sigma(\mathbf{w}_1^T \mathbf{x})] + (1-\theta) \log[\sigma(\mathbf{w}_2^T \mathbf{x})]] - (1-y^{(i)})[\theta \log[1-\sigma(\mathbf{w}_1^T \mathbf{x})] + (1-\theta) \log[1-\sigma(\mathbf{w}_2^T \mathbf{x})]]\}$$

Define $B \triangleq (1 + e^{-\mathbf{w}_1^T \mathbf{x}})^\theta (1 + e^{-\mathbf{w}_2^T \mathbf{x}})^{1-\theta}$,

2

$$= \sum_{i=1}^{n} \{y^{(i)} \log B + (1 - y^{(i)})(\mathbf{w}^{*T}\mathbf{x} + \log B)\}$$

$$= \sum_{i=1}^{n} \{(1 - y^{(i)})\mathbf{w}^{*T}\mathbf{x} + \log B)\}$$

So $RHS - LHS = n(\log B - \log A)$,

$$\log B - \log A = \log \frac{(1 + e^{-\mathbf{w}_1^T\mathbf{x}})^{\theta}(1 + e^{-\mathbf{w}_2^T\mathbf{x}})^{1-\theta}}{1 + e^{-\theta\mathbf{w}_1^T\mathbf{x}} * e^{-(1-\theta)\mathbf{w}_2^T\mathbf{x}}}$$

Because for $\theta \in [0, 1]$, $(1+x_1)^{\theta}(1+x_2)^{1-\theta} \geq 1 + x_1^{\theta}x_2^{1-\theta}$, so $(1 + e^{-\mathbf{w}_1^T\mathbf{x}})^{\theta}(1 + e^{-\mathbf{w}_2^T\mathbf{x}})^{1-\theta} \geq 1 + e^{-\theta\mathbf{w}_1^T\mathbf{x}} * e^{-(1-\theta)\mathbf{w}_2^T\mathbf{x}}$, $RHS - LHS = n(\log B - \log A) \geq 0$.

So for any $\mathbf{w}_1$ and $\mathbf{w}_2$ and $\theta \in [0, 1]$, $L(\theta\mathbf{w}_1 + (1 - \theta)\mathbf{w}_2) \leq \theta L(\mathbf{w}_1) + (1 - \theta)L(\mathbf{w}_2)$. $L_{\mathbf{w}}$ is convex in $\mathbf{w}$. $\square$

(d)

3 Boosting

(a)

(b)

(c)