475 HW 3 Yuanwen Liu

ywu9096

04/29/2020

6.5.

$$g(\tilde{w}) = -\frac{1}{P}\sum_{p=1}^{P}\left[y_p \log(\sigma(\tilde{x}_p^T\tilde{w})) + (1-y_p)\log(1-\sigma(\tilde{x}_p^T\tilde{w})) \right]$$

$$\nabla g(\tilde{w}) = -\frac{1}{P}\sum_{p=1}^{P}\left[y_p \cdot \frac{\sigma'(\tilde{x}_p^T\tilde{w})}{\sigma(\tilde{x}^T\tilde{w})} + (1-y_p)\cdot\frac{-\sigma'(\tilde{x}_p^T\tilde{w})}{1-\sigma(\tilde{x}^T\tilde{w})} \right]\tilde{x}_p$$

$$\nabla g(\tilde{w}) = -\frac{1}{P}\sum_{p=1}^{P}\left[y_p \cdot \frac{\sigma(\tilde{x}_p^T\tilde{w})[1-\sigma(\tilde{x}_p^T\tilde{w})]}{\sigma(\tilde{x}^T\tilde{w})} + (1-y_p)\cdot\frac{-\sigma(\tilde{x}_p^T\tilde{w})[1-\sigma(\tilde{x}_p^T\tilde{w})]}{1-\sigma(\tilde{x}^T\tilde{w})} \right]\tilde{x}_p$$

$$\nabla g(\tilde{w}) = -\frac{1}{P}\sum_{p=1}^{P}\left[y_p[1-\sigma(\tilde{x}_p^T\tilde{w})] + (1-y_p)\cdot(-\sigma(\tilde{x}_p^T\tilde{w})) \right]\tilde{x}_p$$

$$= -\frac{1}{P}\sum_{p=1}^{P}(y_p - \sigma(\tilde{x}_p^T\tilde{w}))\tilde{x}_p$$

$$\nabla^2 g(\vec{w}) = -\frac{1}{P}\sum_{p=1}^{P} -\sigma(\tilde{x}_p^T\tilde{w})(1-\sigma(\tilde{x}_p^T\tilde{w}))\cdot\tilde{x}_p\cdot\tilde{x}_p^T$$

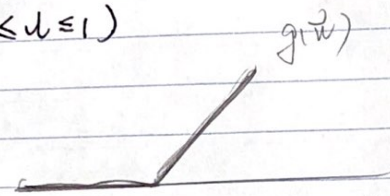
$$= \frac{1}{P}\sum_{p=1}^{P}\sigma(\tilde{x}_p^T\tilde{w})(1-\sigma(\tilde{x}_p^T\tilde{w}))\tilde{x}_p\tilde{x}_p^T$$

6.10

$$g(\vec{w}) = \frac{1}{P}\sum_{p=1}^{P}\max(0, -y_p\tilde{x}_p^T\vec{w})$$

Assuming There are $\vec{w}_1$ and $\vec{w}_2$ corresponding to $g(\vec{w}_1)$, $g(\vec{w}_2)$ respectively.

Choosing an arbitrary point ~~$\lambda g(\vec{w}_1) + (1-\lambda) g(\vec{w}_2)$~~ $\lambda\vec{w}_1 + (1-\lambda)\vec{w}_2$ between $\vec{w}_1$ and $\vec{w}_2$. $(0 \le \lambda \le 1)$



case 1: $g(\vec{w}_1) = g(\vec{w}_2) = 0$

In this case,

$$g(\lambda\vec{w}_1 + (1-\lambda)\vec{w}_2) = \frac{1}{P}\sum_{p=1}^{P}\max(0, -y_p\tilde{x}_p^T(\lambda\vec{w}_1 + (1-\lambda)\vec{w}_2))$$

Since $g(\vec{w}_1) = 0$, $-y_p\tilde{x}_p^T\vec{w}_1 < 0$. Similarly, $-y_p\tilde{x}_p^T\vec{w}_2 < 0$.

Thus, $-y_p\tilde{x}_p^T(\lambda\vec{w}_1 + (1-\lambda)\vec{w}_2) < 0$

$$g(\lambda\vec{w}_1 + (1-\lambda)\vec{w}_2) = 0 = \lambda g(\vec{w}_1) + (1-\lambda) g(\vec{w}_2)$$

Case 2: $g(\vec{w}_1) > 0$, $g(\vec{w}_2) = 0$, or $g(\vec{w}_1) = 0$ $g(\vec{w}_2) > 0$.

Assuming, $g(\vec{w}_1) > 0$, we have, $-y_p\tilde{x}_p^T\vec{w}_1 > 0$

$$g(\lambda\vec{w}_1 + (1-\lambda)\vec{w}_2) = \frac{1}{P}\sum_{p=1}^{P}\max(0, -y_p\tilde{x}_p^T(\lambda\vec{w}_1 + (1-\lambda)\vec{w}_2))$$

Since $-y_p\tilde{x}_p^T\vec{w}_2 \le 0$,

$$-y_p\tilde{x}_p^T\lambda\vec{w}_1 - y_p\tilde{x}_p^T(1-\lambda)\vec{w}_2 \le -y_p\tilde{x}_p^T\lambda\vec{w}_1$$

so $g(\lambda\vec{w}_1 + (1-\lambda)\vec{w}_2) \le \lambda g(\vec{w}_1)$

$$g(\lambda\vec{w}_1 + (1-\lambda)\vec{w}_2) \le \lambda g(\vec{w}_1) + (1-\lambda) g(\vec{w}_2)$$

$g(\vec{w}_2) > 0$ has similar result.

case 3: $g(\vec{w}_1) > 0$ and $g(\vec{w}_2) > 0$.

$$g(\lambda\vec{w}_1 + (1-\lambda)\vec{w}_2) = \lambda g(\vec{w}_1) + (1-\lambda) g(\vec{w}_2).$$

Thus, we have

$g(\lambda\vec{w}_1 + (1-\lambda)\vec{w}_2) \le \lambda g(\vec{w}_1) + (1-\lambda) g(\vec{w}_2)$ for all $0 \le \lambda \le 1$.

Which means this function is convex.