Homework 1

3170102587 蒋仕彪

Problem description

Consider the binary classification problem using Logistic Regression.

Assume I is the set of samples. Denote $\boldsymbol{x}_i$ as the feature vector for i-th sample, and $y_i \in \{-1, 1\}$ is its label.

From Logistic Regression, we want to find a vector $\boldsymbol{a}$ to minimize the function:

$$E(\boldsymbol{a}) = \sum_{i \in I} \ln(1 + e^{-y_i \boldsymbol{a}^T \boldsymbol{x}_i}) \tag{1}$$

Now prove that $E(\boldsymbol{a})$ is a convex function of $\boldsymbol{a}$.

Lemma

Lemma 1: The sum of several convex functions is still convex.

This theory can be derived directly from the definition of convex function.

Lemma 2: If f(x) is twice continuously differentiable, $f(x)$ is convex if and only if $f''(x) \geq 0$.

This theory is classical and can be found in any book about convex optimization, so I skip its proof.

Prove

Because $E(\boldsymbol{a})$ is the sum of n expressions, according to Lemma 1, we can just split them and prove each of them is convex, and finally conclude $E(\boldsymbol{a})$ is convex.

Now we focus on the function $E_i(\boldsymbol{a}) = \ln(1 + e^{-y_i \boldsymbol{a}^T \boldsymbol{x}_i})$.

According to the definition of convex function, $E_i(\boldsymbol{a})$ is the convex function if and only if $\forall 0 \leq \lambda \leq 1$, $E_i(\lambda \boldsymbol{a}_1 + (1-\lambda)\boldsymbol{a}_2) \leq \lambda E_i(\boldsymbol{a}_1) + (1-\lambda) E_i(\boldsymbol{a}_2)$

Note that for each fixed i, $\boldsymbol{x}_i$ remains same. Just set $\theta_k = \boldsymbol{a}_k^T \boldsymbol{x}_i$ and $J(\theta) = \ln(1 + e^{-y_i \theta})$, then the above inequality is **equivalent** to $\forall 0 \leq \lambda \leq 1$, $J(\lambda \theta_1 + (1-\lambda)\theta_2) \leq \lambda J(\theta_1) + (1-\lambda) J(\theta_2)$.

So it's equivalent to prove the following function $J(\theta)$ is the convex function:

$$J(\theta) = \ln(1 + e^{-y_i \theta}) \tag{2}$$

Now we differentiate $J(\theta)$:

1. $J'(\theta) = -\frac{y_i}{e^{y_i \theta} + 1}$
2. $J''(\theta) = \frac{e^{y_i \theta}}{(e^{y_i \theta} + 1)^2}$. Since $y_i \in \{1, -1\}$, thus $J''(\theta) = \frac{e^{\theta}}{(e^{\theta}+1)^2} \geq 0$

According to lemma 2, since $J''(\theta) \geq 0$ for any possible θ, so $J(\theta)$ is the convex function.

Then we know $E_i(\boldsymbol{a})$ is convex of $\boldsymbol{a}$ and finally $E(\boldsymbol{a})$ is convex of $\boldsymbol{a}$, too.