Advanced Machine Learning Midterm Assignment

情報工学系 19B39038 三嶋 隆史

2022 年 7 月 29 日

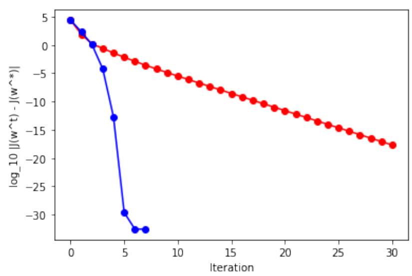
問題 1

1, 2

実装したソースコードを以下のリンクに示す．

3

ニュートン法を実行した結果を青，勾配降下法を実行した結果を赤として，以下の図に示す．結果からわかるように，ニュートン法は初期で収束していることが確認できる．



4

パラメータ w の時に，データ $\boldsymbol{x}$ が y_i に分類される確率 $p(y|\boldsymbol{x},\boldsymbol{w})$ を以下のように定める．

$$p(y_i|\boldsymbol{x},\boldsymbol{w}) = \frac{\exp(\boldsymbol{w}_y^T\boldsymbol{x})}{\sum_{c=1}^{C}\exp(\boldsymbol{w}_c^T\boldsymbol{x})}$$

ここで，データ $\boldsymbol{x}$ に対する損失関数 $l(f(\boldsymbol{x}),y)$ を以下のように定める．

$$l(f(\boldsymbol{x}),y) = -\boldsymbol{w}_y^T\boldsymbol{x} + \ln\left(\sum_{c=1}^{C}\exp(\boldsymbol{w}_c^T\boldsymbol{x})\right)$$

次に $\nabla_{\boldsymbol{w}_y} l$ および $\nabla^2_{\boldsymbol{w}_y} l$ を計算する．

$$\nabla_{\boldsymbol{w}_y} l = -\boldsymbol{x} + \frac{\exp(\boldsymbol{w}_y^T \boldsymbol{x})}{\sum_{i=1}^{C} \exp(\boldsymbol{w}_c^T \boldsymbol{x})} \boldsymbol{x}$$

$$\nabla^2_{\boldsymbol{w}_y} l = \frac{\sum_{i=1, c\neq y}^{C} \exp(\boldsymbol{w}_y^T \boldsymbol{x}) \exp(\boldsymbol{w}_c^T \boldsymbol{x})}{\left\{\sum_{i=1}^{C} \exp(\boldsymbol{w}_c^T \boldsymbol{x})\right\}^2} \boldsymbol{x}\boldsymbol{x}^T$$

これらを使って，勾配降下法とニュートン法を実装する．実装を完了できませんでした．

問題 9

1

この問題は以下のように max と平方根を含まない形に変換できる．

$$\begin{aligned} \text{Minimize :} \quad & z \\ \text{Subject to :} \quad & \forall i, (x_i - x)^2 + (y_i - y)^2 \leq z \\ & z \geq 0 \end{aligned}$$

ここで，ラグランジュ緩和問題を求めると，

$$\begin{aligned} \text{Minimize :} \quad & z - \sum_{i=1}^{n} \alpha_i \left\{ z - (x_i - x)^2 - (y_i - y)^2 \right\} \\ & = \left(1 - \sum_{i=1}^{n} \alpha_i \right) z + \sum_{i=1}^{n} \alpha_i \left\{ (x_i - x)^2 + (y_i - y)^2 \right\} \\ \text{Subject to :} \quad & z \geq 0 \\ & \forall i, \alpha_i \geq 0 \end{aligned}$$

となるから，ここから双対問題に変換すると，

$$\begin{aligned} \text{Maximize :} \quad & \sum_{i=1}^{n} \alpha_i \left\{ (x_i - x)^2 + (y_i - y)^2 \right\} \\ \text{Subject to :} \quad & \sum_{i=1}^{n} \alpha_i \leq 1 \\ & \forall i, \alpha_i \geq 0 \end{aligned}$$

が得られる．

問題 6

$f(\hat{\boldsymbol{w}}) - f(\boldsymbol{w}^*) \leq \epsilon$ となる T を求めるためには，$f(\boldsymbol{w}^{(T)}) - f(\boldsymbol{w}^*) \leq \epsilon$ となる最も小さい $T = T^*$ を見つければ良い．T^* について以下の式が成り立つ．

$$\begin{aligned}
f(\boldsymbol{w}^{(0)}) - f(\boldsymbol{w}^*) &= \left| f(\boldsymbol{w}^{(0)}) - f(\boldsymbol{w}^*) \right| \\
&\leq \sum_{i=0}^{T^*-1} \left| f(\boldsymbol{w}^{(i)}) - f(\boldsymbol{w}^{(i+1)}) \right| + \left| f(\boldsymbol{w}^{(T^*)}) - f(\boldsymbol{w}^*) \right| \\
&\leq \sum_{i=0}^{T^*-1} \left| f(\boldsymbol{w}^{(i)}) - f(\boldsymbol{w}^{(i+1)}) \right| + \epsilon \\
&\leq \sum_{i=0}^{T^*-1} L \left\| \boldsymbol{w}^{(i)} - \boldsymbol{w}^{(i+1)} \right\| + \epsilon \\
&\leq \sum_{i=0}^{T^*-1} \eta L \left\| \nabla f|_{\boldsymbol{w}=\boldsymbol{w}^{(t)}} \right\| + \epsilon \\
&\leq \eta L T^* \max_{0 \leq t < T^*} \left\| \nabla f|_{\boldsymbol{w}=w^{(t)}} \right\| + \epsilon \\
&\leq \frac{\epsilon}{L} \max_{0 \leq t < T^*} \left\| \nabla f|_{\boldsymbol{w}=w^{(t)}} \right\| + \epsilon
\end{aligned}$$

よって，

$$T^* \geq \frac{L(f(\boldsymbol{w}^{(0)}) - f(\boldsymbol{w}^*) - \epsilon)}{\epsilon \max_{0 \leq t < T^*} \left\| \nabla f|_{\boldsymbol{w}=w^{(t)}} \right\|}$$