Hao Zhang haomoodzhang@gmail.com

南京大學

- How to compute dist. mat. w/o any exciplit for loop?

Sol: notations:

- training set $\{(\vec{x}^{(i)}, y^{(i)})\}_{i=1}^{m}$ $\vec{x}^{(i)} \in \mathbb{R}^n, \forall i$
- test set $\{(\vec{x}_{test}^{(i)}, y^{(i)})\}_{i=1}^{m_{test}}$ $\vec{x}_{test}^{(i)} \in \mathbb{R}^n, \forall i$
- design mat. for training set and test set

$$X = \begin{bmatrix} \vec{x}^{(1)T} \\ \vec{x}^{(2)T} \\ \vdots \\ \vec{x}^{(m)T} \end{bmatrix} \in \mathbb{R}^{m \times n}, \quad X_{test} = \begin{bmatrix} \vec{x}_{test}^{(1)T} \\ \vec{x}_{test}^{(2)T} \\ \vdots \\ \vec{x}_{test}^{(m_{test})} \end{bmatrix} \in \mathbb{R}^{m_{test} \times n}$$

- dist. mat. $D \in \mathbb{R}^{m_{test} \times m}$, s.t. $D(i,j) = dist(\vec{x}_{test}^{(i)}, \vec{x}^{(j)})^2$

note that we have made it square to make computation easier

methods:

$$\begin{aligned} D(i,j) &= dist(\vec{x}_{test}^{(i)}, \vec{x}^{(j)})^2 \\ &= \|\vec{x}_{test}^{(i)} - \vec{x}^{(j)}\|^2 \\ &= (\vec{x}_{test}^{(i)} - \vec{x}^{(j)})^T (\vec{x}_{test}^{(i)} - \vec{x}^{(j)}) \\ &= \vec{x}_{test}^{(i)T} \vec{x}_{test}^{(i)} + \vec{x}^{(j)T} \vec{x}^{(j)} - 2 \vec{x}_{test}^{(i)T} \vec{x}^{(j)} \end{aligned}$$

$$\therefore D = \begin{bmatrix} \vec{x}_{test}^{(1)T} \\ \vdots \\ \vec{x}_{test}^{(m)T} \end{bmatrix}^{**2} \vec{1}_n \vec{1}_m^T + \vec{1}_{m_{test}} \vec{1}_n^T \left[\vec{x}^{(1)} \, \vec{x}^{(2)} \cdots \vec{x}^{(m)} \right]$$

$$- 2 \begin{bmatrix} \vec{x}_{test}^{(1)T} \\ \vdots \\ \vec{x}_{test}^{(m)T} \end{bmatrix} \left[\vec{x}^{(1)} \, \vec{x}^{(2)} \cdots \vec{x}^{(m)} \right]$$

$$= X_{test}{**2} \cdot 1_{n \times m} + 1_{m_{test} \times n} \cdot X^{T**2} \cancel{-2X_{test}} - 2 X_{test} X^T$$

- Compute grad. wrt SVM loss

Sol: loss function $J = \frac{1}{m}\sum_{i=1}^{m}\sum_{j\neq y^{(i)}} \max(0, s_j^{(i)} - s_{y^{(i)}}^{(i)} + 1) + \frac{\lambda}{2}\|W\|_F^2$

$$\vec{s}^{(i)} = W\vec{x}^{(i)} + \vec{b}$$

$$\frac{\partial J}{\partial s_j^{(i)}} = \frac{1}{m}\frac{\partial}{\partial s_j^{(i)}} \max(0, s_j^{(i)} - s_{y^{(i)}}^{(i)} + 1) \qquad j \neq y^{(i)}$$

$$= \frac{1}{m} 1\{s_j^{(i)} - s_{y^{(i)}}^{(i)} + 1\}$$

$$\frac{\partial J}{\partial s_{y^{(i)}}^{(i)}} = \frac{1}{m}\sum_{j\neq y^{(i)}} \frac{\partial}{\partial s_{y^{(i)}}^{(i)}} \max(0, s_j^{(i)} - s_{y^{(i)}}^{(i)} + 1)$$

$$= -\frac{1}{m}\sum_{j\neq y^{(i)}} 1\{s_j^{(i)} - s_{y^{(i)}}^{(i)} + 1\}$$

- Compute grad. wrt. softmax loss

Sol: $J = -\frac{1}{m}\sum_{i=1}^{m}\log\frac{\exp(s_{y^{(i)}}^{(i)})}{\sum_{j=1}^{K}\exp(s_j^{(i)})} + \frac{\lambda}{2}\|W\|_F^2$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left(s_{y^{(i)}}^{(i)} - \log\left(\sum_j \exp(s_j^{(i)})\right)\right) + \frac{\lambda}{2}\|W\|_F^2$$

$$= -\frac{1}{m}\sum_{i=1}^{m} s_{y^{(i)}}^{(i)} + \frac{1}{m}\sum_{i=1}^{m}\log\left(\sum_j \exp(s_j^{(i)})\right) + \frac{\lambda}{2}\|W\|_F^2$$

$$\frac{\partial J}{\partial s_j^{(i)}} = \frac{1}{m}\frac{\partial}{\partial s_j^{(i)}}\log\left(\sum_j \exp(s_j^{(i)})\right) \qquad j \neq y^{(i)}$$

$$= \frac{1}{m}\frac{\exp(s_j^{(i)})}{\sum_k \exp(s_k^{(i)})}$$

$$\frac{\partial J}{\partial s_{y^{(i)}}^{(i)}} = \frac{1}{m}\frac{\exp(s_{y^{(i)}}^{(i)})}{\sum_k \exp(s_k^{(i)})} - \frac{1}{m}$$