Support Vector Machine Vectorized proof

Remon Kamal

March 20, 2020

Abstract

This is the proof of the vectorized form of SVM, I do this with lots of Matrix Math, I hope you like my latex skills

1 proof

$$loss = \frac{1}{N}\sum_{i=0}^{N-1} loss_i + \lambda \left\| W \right\|_2 \tag{1}$$

$$loss_i = \sum_{k=0}^{C-1} T_{ik} \tag{2}$$

$$T_{ik} = \begin{cases} 0 & , margin_{ik} \leq 0 \\ S_{ik} - S_{iy_j} + 1 & , margin_{ik} > 0 \end{cases} \tag{3}$$

$$\tag{4}$$

Where

$$margin_{ik} = S_{ik} - S_{iy_j} + 1 \tag{5}$$

$$\tag{6}$$

And

$$S_{ik} = X_i W_k \tag{7}$$

$$W_k \rightarrow \text{k-Coloum vector of W matrix} \tag{8}$$

$$W_k = W u_k \tag{9}$$

$$u_k \rightarrow \text{Unit coloum vector in direction of k} \tag{10}$$

$$S_{ik} = X_i W u_k \tag{11}$$

$$\nabla_W loss = \frac{1}{N} \nabla_W loss_i + \lambda \nabla_W ||W||_2^2 \tag{12}$$

$$\nabla_W ||W||_2^2 = 2W \tag{13}$$

$$\nabla_W T_{ik} = \begin{cases} 0 & , margin_{ik} \leq 0 \\ \nabla_W S_{ik} - \nabla_W S_{iy_j} & , margin_{ik} > 0 \end{cases} \tag{14}$$

$$(\nabla_W)_{lm} = \frac{\partial}{\partial W_{lm}} \tag{15}$$

$$\begin{aligned} (\nabla_W S_{ik})_{lm} &= \frac{\partial}{\partial W_{lm}} (X_i W u_k) \\ &= X_i \frac{\partial}{\partial W_{lm}} (W u_k) \\ &= X_i \frac{\partial}{\partial W_{lm}} \left(u_k^T W^T\right) \\ &= X_i u_k^T \frac{\partial}{\partial W_{lm}} \left(W^T\right) \end{aligned} \tag{16}$$

$$\frac{\partial}{\partial W_{lm}} \left(W^T\right) = \omega_{lm}^T \tag{17}$$

Where ω_{lm} is a matrix full of zeros except at $index_{lm}$ which is 1 from ... in ...

$$X_i u_k^T \omega_{lm}^T = X_i \omega_{lm} u_k \tag{18}$$

$$\omega_{lm} u_k = \begin{cases} 0 & k \neq l \\ u_m & k = l \end{cases} \tag{19}$$

$$X_i \omega_{lm} u_k = \begin{cases} 0 & k \neq l \\ X_i u_m = X_{im} & k = l \end{cases} \tag{20}$$

Where X_{im} is m^{th} element in X_i

$$\text{So } \nabla_W (S_{ik}) = \begin{bmatrix} 0 & 0 & | & 0 \\ 0 & 0 & X_i & 0 \\ 0 & 0 & | & 0 \\ 0 & 0 & k_{th} - col & 0 \end{bmatrix} \tag{21}$$

$$\text{Simlarly } \nabla_W \left(S_{y_j}\right) = \begin{bmatrix} 0 & 0 & | & 0 \\ 0 & 0 & X_i & 0 \\ 0 & 0 & | & 0 \\ 0 & 0 & y_{j-th} col & 0 \end{bmatrix} \tag{22}$$

$$\text{So } \nabla_W T_{ik} = \begin{cases} \begin{bmatrix} | & 0 & | & 0 \\ X_i & 0 & X_i & 0 \\ | & 0 & | & 0 \\ k_{th} - col & 0 & y_{j-th} - col & 0 \end{bmatrix} & , margin_{ik} \leq 0 \\ \text{zero-matrix} & , margin_{ik} > 0 \end{cases} \tag{23}$$

$$\nabla_W loss_i = \begin{bmatrix} | & 0 & | & 0 \\ X_i \text{ if } morgin_{ik} > 0 \text{ else zeros} & 0 & -m_i X_i & 0 \\ | & 0 & | & 0 \\ k_{th} - col & 0 & y_{j-th} - col & 0 \end{bmatrix} \tag{24}$$

Where m_i is the number of $margin_{ik}$ in i$^{\text{th}}$ example that is greater than zero, So we can assume C the cost matrix where C_{ic} corresponds to the i$^{\text{th}}$ example and c$^{\text{th}}$ class cost where

$$C_{in} = \begin{cases} -m_i & n = y_j \text{ n is the correct label for } example_i \\ 1 & margin_{in} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

$$\left(\sum_i^N \nabla_W loss_i \right)_{lm} = \sum_i^N C_{im} X_{il} \tag{26}$$

$$\tag{27}$$

$$\text{So } \sum_i^N \nabla_W loss_i = X^T C \tag{28}$$

$$\text{So the final loss expression } \nabla_W loss = \frac{1}{N} \left(X^T C \right) + 2\lambda W \tag{29}$$

How about this (1) Equation label