Support Vector Machine Vectorized proof

Remon Kamal

March 21, 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

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

Where N is number of examples, W is the weight matrix and λ is the regression weight

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

Where $loss_i$ is the loss per example, T_{ik} is loss of class k in example i

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

Where $margin_{ik}$ is the margin loss of class k in example i, and it is given by

$$margin_{ik} = S_{ik} - S_{iy_j} + 1 \tag{4}$$

And S_{ik} is the score of class k in example i

$$S_{ik} = X_i W_k \tag{5}$$

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

$$W_k = W u_k \tag{6}$$

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

Substituting From (6) in (5) we get:

$$S_{ik} = X_i W u_k \tag{7}$$

Taking the ∇_W of (1) we have

$$\nabla_W loss = \frac{1}{N} \nabla_W loss_i + \lambda \nabla_W ||W||_2^2 \tag{8}$$

We divide the left part of (8) and process each part separately

$$\nabla_W ||W||_2^2 = 2W \tag{9}$$

$$\nabla_W loss_i = \sum_{k=0}^{c-1} \nabla_W T_{ik} \tag{10}$$

where c is the number of classes in our problem

$$\nabla_W T_{ik} = \begin{cases} 0 & , margin_{ik} \leq 0 \\ \nabla_W S_{ik} - \nabla_W S_{iy_j} & , margin_{ik} > 0 \end{cases} \tag{11}$$

$$(\nabla_W)_{lm} = \frac{\partial}{\partial W_{lm}} \tag{12}$$

From equation (7) and substituting in equation (12) we get:

$$\begin{aligned} (\nabla_W S_{ik})_{lm} &= \frac{\partial}{\partial W_{lm}} (X_i W u_k) \\ &= X_i \frac{\partial}{\partial W_{lm}} (W u_k) \\ &= X_i \frac{\partial}{\partial W_{lm}} \left(u_k^T W^T\right) \\ &= X_i u_k^T \frac{\partial}{\partial W_{lm}} \left(W^T\right) \end{aligned} \tag{13}$$

$$\frac{\partial}{\partial W_{lm}} \left(W^T\right) = \omega_{lm}^T \tag{14}$$

Where ω_{lm} is a matrix full of zeros except at $index_{lm}$ which is 1 By substitution from equation (14) in equation (13) we get:

$$\begin{aligned} (\nabla_W S_{ik})_{lm} &= X_i u_k^T \omega_{lm}^T \\ &= X_i \omega_{lm} u_k \end{aligned} \tag{15}$$

$$\omega_{lm} u_k = \begin{cases} 0 & k \neq l \\ u_m & k = l \end{cases} \tag{16}$$

So by getting the right hand side from equation (15) and substituting by equation (16) we get:

$$X_i \omega_{lm} u_k = \begin{cases} 0 & k \neq l \\ X_i u_m = X_{im} & k = l \end{cases} \tag{17}$$

Where X_{im} is m$^{\text{th}}$ element in X_i
So from equation (15) and equation (17) we get:

$$\nabla_W \left(S_{ik}\right) = \begin{bmatrix} 0 & 0 & | & 0 \\ 0 & 0 & X_i & 0 \\ 0 & 0 & | & 0 \\ 0 & 0 & k_{th} - col & 0 \end{bmatrix} \tag{18}$$

$$\text{Simlarly } \nabla_W \left(S_{y_j}\right) = \begin{bmatrix} 0 & 0 & | & 0 \\ 0 & 0 & X_i & 0 \\ 0 & 0 & | & 0 \\ 0 & 0 & y_{j-th} col & 0 \end{bmatrix} \tag{19}$$

By substituting from equation (18) and (19) in equation (3) we get:

$$\text{So } \nabla_W T_{ik} = \begin{cases} \begin{bmatrix} | & 0 & | & 0 \\ X_i & 0 & X_i & 0 \\ | & 0 & | & 0 \\ k_{th} - col & 0 & y_{j-th} - col & 0 \end{bmatrix} & , margin_{ik} \leq 0 \\ \text{zero-matrix} & , margin_{ik} > 0 \end{cases} \tag{20}$$

From equation (20) in (2) we get:

$$\nabla_W loss_i = \begin{bmatrix} | & 0 & | & 0 \\ X_i \text{ if } morgin_{ik} > 0 \text{ else zeros} & 0 & -m_i X_i & 0 \\ | & 0 & | & 0 \\ k_{th} - col & 0 & y_{j-th} - col & 0 \end{bmatrix} \tag{21}$$

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_{in} corresponds to the i$^{\text{th}}$ example and c$^{\text{th}}$ class cost where

$$C_{in} = \begin{cases} -m_i & n = y_j \text{ n is the correct label for } example_i \\ 1 & margin_{in} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

And we get:

$$\left(\sum_{i}^{N} \nabla_W loss_i\right)_{lm} = \sum_{i}^{N} C_{im} X_{il} \tag{23}$$

$$\text{So } \sum_{i}^{N} \nabla_W loss_i = X^T C \tag{24}$$

And from (24) and (9) and substituting in (8) we get the final vectorized loss expression:

$$\nabla_W loss = \frac{1}{N}\left(X^T C\right) + 2\lambda W \tag{25}$$