One of the main Drawbacks of K-NN is the curse of Dimensionality. As the feature space gets larger, the feature vectors become sparser and the feature vectors become sparser and as a result distance between them as a result distance can be dominated increases. and distance can be dominated increases. and distance can be dominated increases. As a result, by irrelevant attributes. As a result, the neighborhood of a test instance the neighborhood of a test instance doesn't contain instances that one Joinilar' to it.

In linear classification, we use a   
Qinear scoring function
$$f(x^{(i)}, W, b) = W x^{(i)} + b$$

where,  

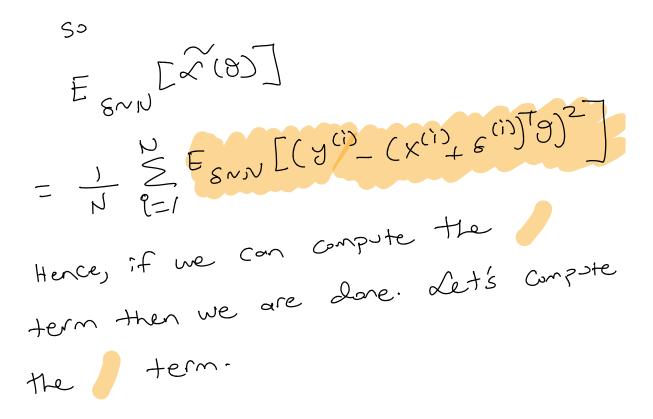
$$W = \begin{bmatrix} -w_1^T \\ -w_2^T \\ \vdots \\ -w_c^T - \end{bmatrix}$$
  
where C is the number of classes. The  
yth entry of  $f(x^{(i)}, W, b)$  is the  
jth entry of  $f(x^{(i)}, W, b)$  is the  
confidence score of image  $x^{(i)}$  belonging  
confidence score of image  $x^{(i)}$  belonging  
to class j. W and b one parameters  
to class j. W and b one parameters  
of the scoring function.

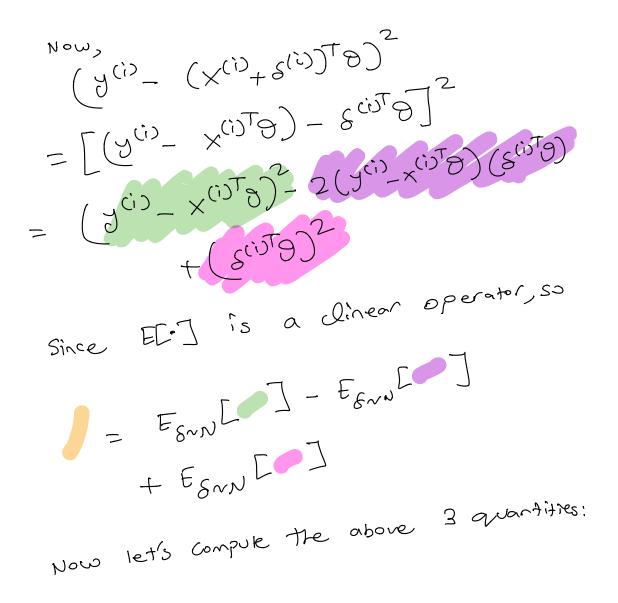
Softmax classifier:  
In softmax classifier we view the  
scores as unnormalized log  
Probabilities for each class and use  
a cross-entropy loss of the form  

$$Li(\Theta) = -\log\left(\frac{e^{fy}}{E^{-}e^{fy}}\right)$$
  
where  $fy_{e} = wit x^{(i)} + bi$   
 $fy_{0} = w_{0}^{T} x^{(i)} + bj$   
Then the loss for softmax is  
 $L(\Theta) = -\log\left(\frac{e^{fy}}{E^{-}e^{fy}}\right)$ 

$$\frac{3}{\chi(\theta)} = \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - (x^{(i)} + \varepsilon^{(i)})^{T}g)^{2}$$

a) since E[.] is a linear operator,



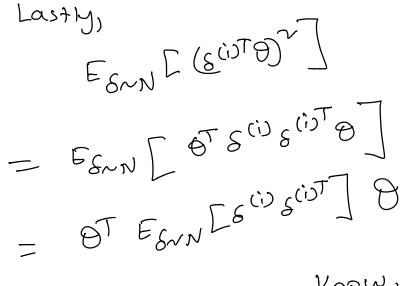


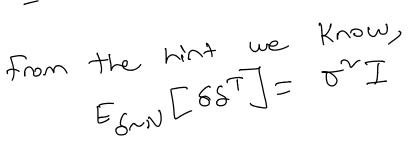
since has no 
$$\delta$$
 dependence, so  
 $E_{SVN}EJ = (y^{(i)} - x^{(i)}T\Theta)^2$ 

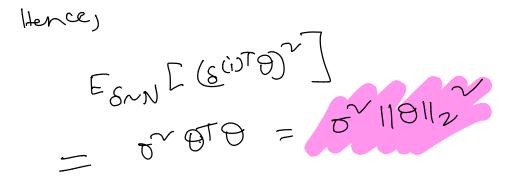
Now,  

$$E_{SNN} \left[ -2(y^{(i)} - x^{(i)T} \partial) (S^{(i)T} \partial) \right]$$
  
 $= -2(y^{(i)} - x^{(i)T} \partial) E_{SNN} \left[ S^{(i)T} \partial \right]$   
 $= 0 \in \mathbb{R}^{d}$   
From problem statement,  $E[S^{(i)}] = 0 \in \mathbb{R}^{d}$ 

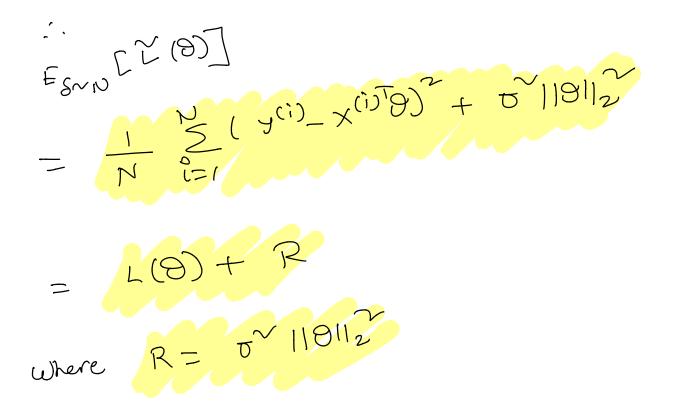
Hence,  $E_{SNN} \left[ -2(y^{(i)} - x^{(i)} - y^{(i)}) \right]$ = 0



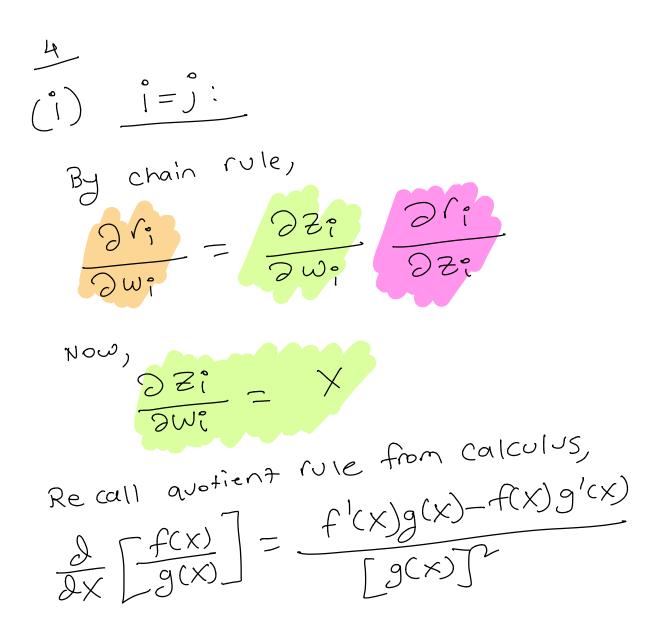


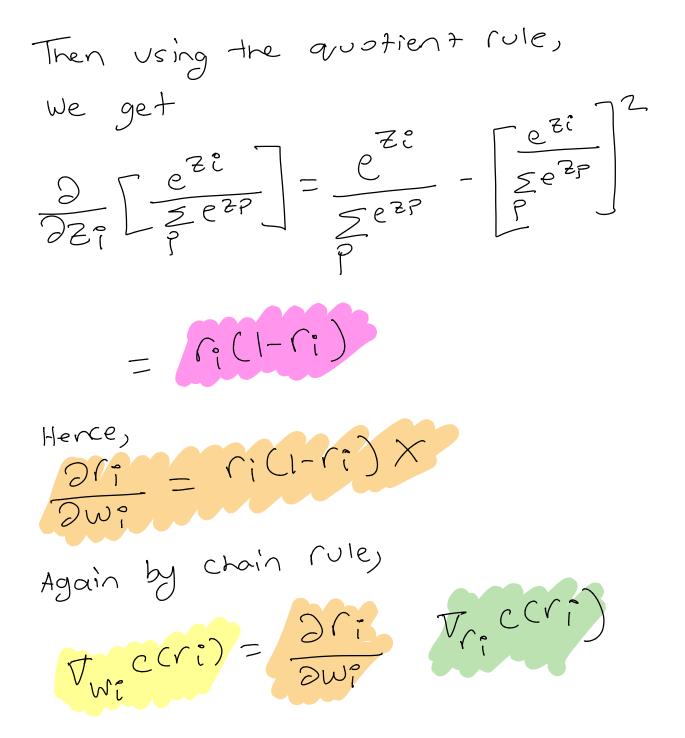


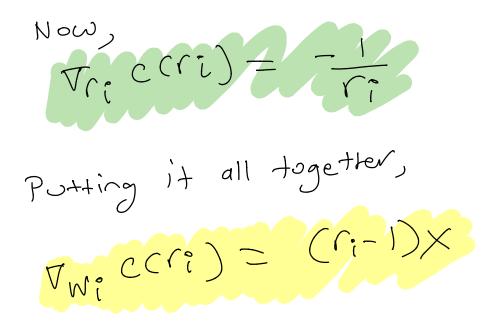
Putting it all together,  
$$= (y^{(i)} - x^{(i)T}\theta)^2 + \sigma^2 ||\theta||_2^2$$

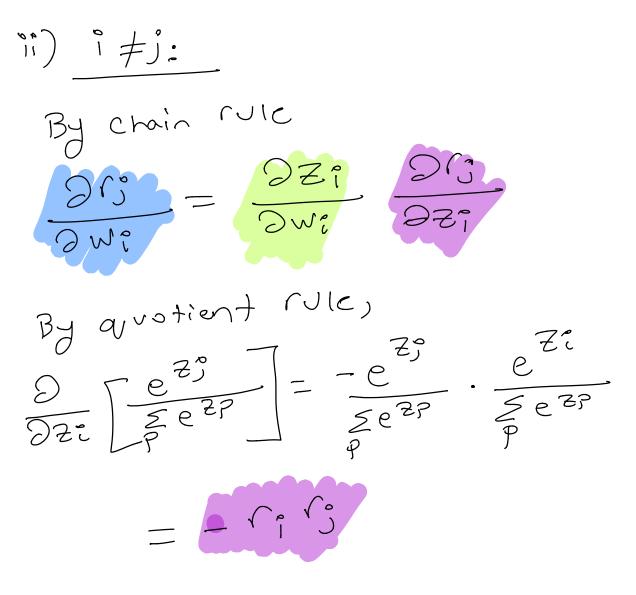


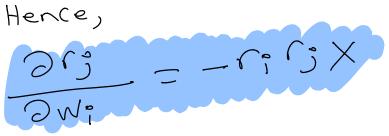
c) As the regularization strength T=20, then the objective of the Cost Function is to minimize the L-2 norm of Parameters & and hence Ø->0 and the model will underfit the data.







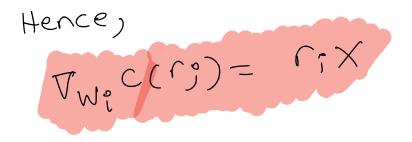


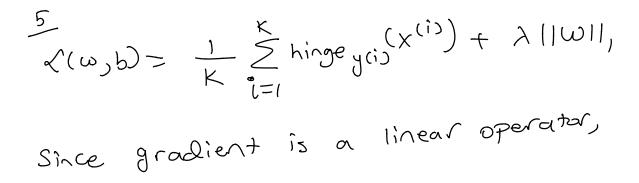


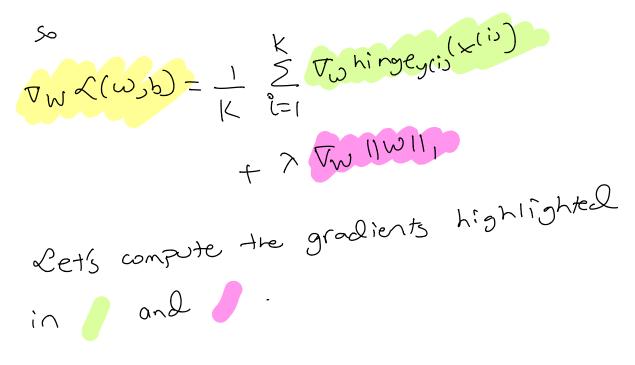
Again by chain rule,  

$$\nabla_{W_1} c(r_3) = \frac{\partial r_3}{\partial W_1} \quad \nabla_{r_3} c(r_3)$$

$$\frac{Now}{\nabla r_{3}^{2}C(r_{3}^{2})} = -\frac{1}{r_{3}^{2}}$$







$$F_{W} = \frac{1}{100} \left( x^{(i)} \right) = \max(0, 1 - y^{(i)}(w^{T}x^{(i)}+b) \right)$$
The above function is a Piecewise Dinear function and can be represented as
$$\int (y^{(i)}(w^{T}x^{(i)}+b) \right) + \int (w^{T}x^{(i)}+b) \right)$$

$$\int (y^{(i)}(w^{T}x^{(i)}+b) - \int (w^{T}x^{(i)}+b) - \int (w^$$

If 
$$y^{(i)}(\omega^{T}x^{(i)}+b) < 1$$
, then  
 $\nabla_{w}$  hirdle  $y^{(i)}(x^{(i)}) = -y^{(i)}x^{(i)} \in \mathbb{R}^{d}$   
Potting it all together,  
 $\nabla_{w}$  hirdle  $y^{(i)}(x^{(i)})$   
 $= II _{2} y^{(i)}(\omega^{T}x^{(i)}+b) < 12$ 

where,  

$$I = y^{(i)} (w^{T} x^{(i)} + b) < 13^{i}$$
  
vector of all ones if  $y^{(i)} (w^{T} x^{(i)} + b) < 1$   
or a vector of all zeros if  
 $y^{(i)} (w^{T} x^{(i)} + b) > 1$ 

 $\frac{||\omega||}{\sum_{i=1}^{a} |\omega_i|}$ 

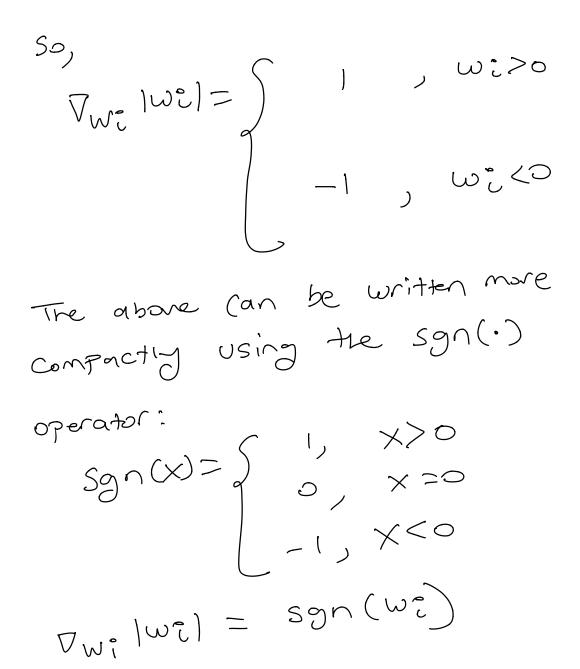
•

Hence,  

$$\nabla_{W} ||w||_{1} = \nabla_{W_{2}} |w_{2}|$$
  
 $\vdots$   
 $\nabla_{W_{2}} |w_{2}|$   
 $\nabla_{W_{2}} |w_{2}|$ 

Now,  

$$|w_{l}| = \int w_{l}, w_{l} > 0$$
  
 $|w_{l}| = \int w_{l}, w_{l} < 0$   
 $-w_{l}, w_{l} < 0$   
for  $i = 1, \dots, Q$ .



Hence,  

$$\overline{\mathcal{T}_{W}} || ||_{I} = Sgn(w) \in \mathbb{R}^{d}$$
  
where  $Sgn(\cdot)$  operates un each  
element of  $W$ .  
Rutting everything together,  
 $\overline{\mathcal{T}_{W}} \propto (w_{3}b)$   
 $= \frac{1}{K} \sum_{i=1}^{K} \overline{\mathbb{T}}_{2y}^{(i)} (w_{1}x^{(i)}+b) < 13 \xrightarrow{(i)}{(1 + 1)} < 13 \xrightarrow{(i)}{(1 + 1)}$