Middle East Technical University | Department of Computer Engineering

CENG 499

Intro to ML

Fall 2024-2025

Assignment 1

Name SURNAME: Aly Asad GILANI
Student ID: 2547875

Answer for Part 1: Regression

First layer weight update rule (w_{ij}^{new}):

We need to expand:

$$w_{ij}^{new} = w_{ij} - \alpha \frac{\partial SE(y, O_0)}{\partial w_{ij}}$$

To do this, we need to find $\frac{\partial SE(y,O_0)}{\partial w_{ij}}$. We can use the chain rule:

$$\frac{\partial SE(y, O_0)}{\partial w_{ij}} = \frac{\partial SE(y, O_0)}{\partial O_0} \cdot \frac{\partial O_0}{\partial H_j} \cdot \frac{\partial H_j}{\partial w_{ij}}$$

So, we need to find all three partial derivatives on the right-hand side.

For the first term, from the PDF, we have:

$$SE(y, O_0) = (y - O_0)^2$$

We find the derivative with respect to O_0 using basic differentiation rules:

$$\frac{\partial SE(y, O_0)}{\partial O_0} = -2(y - O_0)$$

For the second term, from the PDF, we have:

$$O_0 = \sum_{j=0}^{3} H_j * \gamma_{j0}$$

We find the derivative with respect to H_j using basic differentiation rules:

$$\frac{\partial O_0}{\partial H_j} = \gamma_{j0}$$

For the third term, from the PDF, we have:

$$H_j = \sigma(\sum_{i=0}^{2} I_i * w_{ij})$$

The derivative of the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$ with respect to x is known as $\sigma(x)(1-\sigma(x))$. Using this and the chain rule, we can differentiate H_j with respect to w_{ij}:

$$\frac{\partial H_j}{\partial w_{ij}} = H_j(1-H_j) * I_i$$

Finally, putting all partial derivatives together, we get the partial derivative of $SE(y, O_0)$ with respect to w_{ij}:

$$\frac{\partial SE(y, O_0)}{\partial w_{ij}} = \frac{\partial SE(y, O_0)}{\partial O_0} \cdot \frac{\partial O_0}{\partial H_j} \cdot \frac{\partial H_j}{\partial w_{ij}}$$

$$\frac{\partial SE(y, O_0)}{\partial w_{ij}} = -2(y - O_0)\gamma_{j0} H_j(1-H_j) I_i$$

Putting this in the original equation, we get the weight update rule:

$$w_{ij}^{new} = w_{ij} - \alpha \frac{\partial SE(y, O_0)}{\partial w_{ij}}$$

$$\boldsymbol{w_{ij}^{new} = w_{ij} + \alpha * 2(y - O_0)\gamma_{j0} H_j(1-H_j) I_i}$$

Since we know I_0 is 1, we have:

$$\boldsymbol{w_{bias}^{new} = w_{bias} + \alpha * 2(y - O_0)\gamma_{j0} H_j(1-H_j)}$$

Second layer weight update rule (γ_{k0}^{new}):

We need to expand:

$$\gamma_{k0}^{new} = \gamma_{k0} - \alpha \frac{\partial SE(y, O_0)}{\partial \gamma_{k0}}$$

To do this, we need to find $\frac{\partial SE(y,O_0)}{\partial \gamma_{k0}}$. We can use the chain rule:

$$\frac{\partial SE(y, O_0)}{\partial \gamma_{k0}} = \frac{\partial SE(y, O_0)}{\partial O_0} \cdot \frac{\partial O_0}{\gamma_{k0}}$$

So, we need to find both partial derivatives on the right-hand side. We already know $\frac{\partial SE(y,O_0)}{\partial O_0}$ from our working above:

$$\frac{\partial SE(y, O_0)}{\partial O_0} = -2(y - O_0)$$

For the second term, from the PDF, we have:

$$O_0 = \sum_{k=0}^{3} H_k * \gamma_{k0}$$

We find the derivative with respect to γ_{k0} using basic differentiation rules:

$$\frac{\partial O_0}{\partial \gamma_{k0}} = H_k$$

Finally, putting both partial derivatives together, we get the partial derivative of $SE(y, O_0)$ with respect to γ_{k0}:

$$\frac{\partial SE(y, O_0)}{\partial \gamma_{k0}} = \frac{\partial SE(y, O_0)}{\partial O_0} \cdot \frac{\partial O_0}{\gamma_{k0}}$$

$$\frac{\partial SE(y, O_0)}{\partial \gamma_{k0}} = -2(y - O_0) * H_k$$

Putting this in the original equation, we get the weight update rule:

$$\gamma_{k0}^{new} = \gamma_{k0} - \alpha \frac{\partial SE(y, O_0)}{\partial \gamma_{k0}}$$

$$\boldsymbol{\gamma_{k0}^{new} = \gamma_{k0} + \alpha * 2(y - O_0) H_k}$$

Since we know H_0 is 1, we have:

$$\boldsymbol{\gamma_{bias}^{new} = \gamma_{bias} + \alpha * 2(y - O_0)}$$

Answer for Part 1: Classification

First layer weight update rule (w_{ij}^{new}):

We need to expand:

$$w_{ij}^{new} = w_{ij} - \alpha \frac{\partial CE([l_0, l_1, l_2], O = [O_0, O_1, O_2])}{\partial w_{ij}} = w_{ij} - \alpha \frac{\partial CE}{\partial w_{ij}}$$

To do this, we need to find $\frac{\partial CE}{\partial w_{ij}}$. We can use the chain rule:

$$\frac{\partial CE}{\partial w_{ij}} = \frac{\partial CE}{\partial O_m} \cdot \frac{\partial O_m}{\partial X_k} \cdot \frac{\partial X_k}{\partial H_j} \cdot \frac{\partial H_j}{\partial w_{ij}}$$

So, we need to find all four partial derivatives on the right-hand side.

For the first term, from the PDF, we have:

$$CE(l, O) = -\sum_{m=0}^{2} l_m * log(O_m)$$

Taking the derivative with respect to O_m:

$$\frac{\partial CE}{\partial O_m} = -\frac{l_m}{O_m}$$

For the second term, from the PDF, we have:

$$O_k = softmax(X_k, X)$$

We know that the derivative of the softmax function is a Jacobian matrix with the following values:

$$\frac{\partial O_m}{\partial X_k} = \begin{cases} O_k(1 - O_k), & \text{if } m = k \\ -O_m O_k, & m \neq k \end{cases}$$

Combining the first and second terms, we can get:

$$\frac{\partial CE}{\partial X_k} = \sum_{m=0}^{2} \frac{\partial CE}{\partial O_m} \cdot \frac{\partial O_m}{\partial X_k}$$

Expanding for when $m = k$ and $m \neq k$:

$$\frac{\partial CE}{\partial X_k} = -\frac{l_k}{O_k} \cdot O_k(1 - O_k) - \sum_{m \neq k} \frac{l_m}{O_m} \cdot (-O_m O_k)$$

Simplifying:

$$\frac{\partial CE}{\partial X_k} = -l_k(1 - O_k) + \sum_{m \neq k} l_m O_k$$

Factoring out O_k from the summation:

$$\frac{\partial CE}{\partial X_k} = -l_k(1 - O_k) + O_k \sum_{m \neq k} l_m$$

Since l is one-hot encoded, we know that $\sum_m l_m = 1$, so $\sum_{m \neq k} l_m = 1 - l_k$. Substituting this into the equation:

$$\frac{\partial CE}{\partial X_k} = -l_k(1 - O_k) + O_k(1 - l_k)$$

Finally, simplifying gives us:

$$\frac{\partial CE}{\partial X_k} = O_k - l_k$$

We have already found the third and fourth term while solving the regression problem, and they are:

$$\frac{\partial X_k}{\partial H_j} = \gamma_{jk}$$

$$\frac{\partial H_j}{\partial w_{ij}} = H_j(1 - H_j) * I_i$$

Combining all terms together, we get:

$$\frac{\partial CE}{\partial w_{ij}} = \frac{\partial CE}{\partial X_k} \cdot \frac{\partial X_k}{\partial H_j} \cdot \frac{\partial H_j}{\partial w_{ij}}$$

$$\frac{\partial CE}{\partial w_{ij}} = (O_k - l_k) * \gamma_{jk} * H_j(1 - H_j) * I_i$$

Finally, we can write the weight update rule by putting this in the original equation:

$$w_{ij}^{new} = w_{ij} - \alpha \frac{\partial CE}{\partial w_{ij}}$$

$$\boldsymbol{w_{ij}^{new} = w_{ij} - \alpha(O_k - l_k) * \gamma_{jk} * H_j(1 - H_j) * I_i}$$

Since we know I_0 is 1, we have:

$$\boldsymbol{w_{bias}^{new} = w_{bias} - \alpha(O_k - l_k) * \gamma_{jk} * H_j(1 - H_j)}$$

Second layer weight update rule (γ_{jk}^{new}):

We need to expand:

$$\gamma_{jk}^{new} = \gamma_{jk} - \alpha \frac{\partial CE([l_0, l_1, l_2], O = [O_0, O_1, O_2])}{\partial \gamma_{jk}} = \gamma_{jk} - \alpha \frac{\partial CE}{\partial \gamma_{jk}}$$

To do this, we need to find $\frac{\partial CE}{\partial \gamma_{jk}}$. We can use the chain rule:

$$\frac{\partial CE}{\partial \gamma_{jk}} = \frac{\partial CE}{\partial X_k} \cdot \frac{\partial X_k}{\partial \gamma_{jk}}$$

So, we need to find both partial derivatives on the right-hand side. We have already found the first term in our working above for the first layer weight update:

$$\frac{\partial CE}{\partial X_k} = O_k - l_k$$

We have also found the second term in our solution for the regression problem:

$$\frac{\partial X_k}{\partial \gamma_{jk}} = H_j$$

Combining both terms together, we get:

$$\frac{\partial CE}{\partial \gamma_{jk}} = \frac{\partial CE}{\partial X_k} \cdot \frac{\partial X_k}{\partial \gamma_{jk}}$$

$$\frac{\partial CE}{\partial \gamma_{jk}} = (O_k - l_k)H_j$$

Putting this in the original equation, we get the weight update rule:

$$\gamma_{jk}^{new} = \gamma_{jk} - \alpha \frac{\partial CE}{\partial \gamma_{jk}}$$

$$\boldsymbol{\gamma_{jk}^{new} = \gamma_{jk} - \alpha(O_k - l_k)H_j}$$

Since we know H_0 is 1, we have:

$$\boldsymbol{\gamma_{bias}^{new} = \gamma_{bias} - \alpha(O_k - l_k)}$$