Task 1

a) $C(w) = \frac{1}{N} \sum_{n=1}^{N} C^n$

$C^n(w) = -(y^n \ln(\hat{y}^n) + (1-y^n)\ln(1-\hat{y}^n))$

$\frac{\partial f(x^n)}{\partial w_i} = x_i^n f(x^n)(1-f(x^n))$

$\frac{\partial C^n(w)}{\partial w_i} = x_i^n (C(w^n)(1-C(w^n)))$

b) $$\frac{\partial C(w)}{\partial w_{kj}} = \frac{\partial C}{\partial a}\frac{\partial a}{\partial z}\frac{\partial z}{\partial w}$$

$$\left.\frac{\partial C}{\partial a}\right|_{a=\hat{y}=f(z)} = \frac{\partial\left(-\sum_{k=1}^{K} y_k \ln(\hat{y}_k)\right)}{\partial \hat{y}_k}$$

$i=j$, $$\frac{\partial \hat{y}_i}{\partial x_k} = \frac{e^{x_i}\sum_k e^{x_k} - e^{x_i}e^{x_i}}{(\sum_k e^{x_k})^2} = \hat{y}_i(1-\hat{y}_i)$$

$i \neq j$, $$\frac{\partial y_i}{\partial x_j} = \frac{0 - e^{x_i}e^{x_j}}{\left(\sum_k e^{x_k}\right)^2} = -\hat{y}_i\hat{y}_j$$

$$\frac{\partial C}{\partial y_i} = -\sum_i y_i \frac{1}{\hat{y}_i}$$

$$\frac{\partial C}{\partial x_j} = -\sum_{i\neq j} y_i \frac{1}{\hat{y}_i}\frac{\partial y_i}{\partial x_j} - y_j\frac{1}{\hat{y}_j}\frac{\partial y_j}{\partial x_j}$$

$$= -\sum_{i\neq j} y_i \frac{1}{\hat{y}_i}(-\hat{y}_i\hat{y}_j) - y_i\frac{1}{\hat{y}_j}\hat{y}_j(1-\hat{y}_j)$$

$$= y_j - y_i$$

$$\frac{\partial C}{\partial a} = -x_j(y_i - \hat{y}_i)$$

4 a)

$$R(w) = \|w\|^2 = \frac{1}{2}\sum_{i,j} w_{i,j}^2$$

$$J = \mathcal{C}(w) + \lambda R(w)$$

$$\frac{\partial J}{\partial w} = \frac{\partial C}{\partial w} + \frac{\partial\left(\lambda \frac{1}{2}\sum_{i,j} w_{i,j}^2\right)}{\partial w} = \frac{\partial C}{\partial w} + \lambda \cdot w$$

~~correct~~ found!

X_train : Images for training

Y_train : labels for training

Image 1 = Image.resize(28, 28)

plt.imshow(Image 1)

$\hat{Y}$: prediction

Y : ground truth

Gradient descent:

$$W_{t+1} = W_t - \alpha \frac{\partial(\hat{C}(w))}{\partial w}$$

Loss plot

Y-axis : avg loss

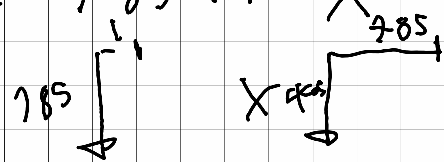
X-axis : Number of testing steps

$X : 4005 \times 785$ →

↑ rader ↑ kolumner

$w : 785 \times 1$ →

$wX = 785 \times 1 \; X \; 4005 \times 785$



Xw går

$w^T X = 1 \times 785 \; X \; 4005 \times 785$