Abhay Doke

Intelligent Visual Computing

Assignment 1

1. **Learned Weights w**
   a. **Without Regularization**

-2.9267 0.1149 0.3401 0.4850 1.6592 7.0114 -7.5944 -15.2727 -4.8174 34.8158 -19.2865

   b. **With Regularization**

-6.4084 -0.0090 -0.0001 0.0060 0.0006 -0.0046 -0.0052 -0.0073 -0.0074 24.9215 -0.0018

2. **Output Probabilities**
   a. **Without Regularization**

Test Example No.	Probability	Test Example No.	Probability
1	0.9999	12	0.9939
2	0.0046	13	0.0001
3	0.0086	14	0.9998
4	0.9997	15	0.9995
5	0.0059	16	0.0607
6	0.0001	17	0.9997
7	0.0035	18	0.9998
8	0.9993	19	0.0133
9	0.9999	20	0.0001
10	0.0002	21	0.0065
11	0.9997	22	0.3719

Test Error = 4.55% (21/22 correct predictions)

   b. **With Regularization**

Test Example No.	Probability	Test Example No.	Probability
1	0.9795	12	0.8841
2	0.0630	13	0.0040
3	0.0787	14	0.9778
4	0.9669	15	0.9586
5	0.19600	16	0.1331
6	0.0077	17	0.9704
7	0.0884	18	0.9732
8	0.9503	19	0.0674
9	0.9860	20	0.0047
10	0.0055	21	0.0653
11	0.9915	22	0.5852

Test Error = 0% (22/22 correct predictions)

E) Derivation of gradient for training

→ With L_1 regularization, we are trying to minimize the loss given by

$$J(W) = L(W) + \lambda \sum_{d} |W_d|$$

$L(W)$ ↓ Cross Entropy Loss

$\lambda \sum_{d} |W_d|$ ↓ Regularization Loss

Gradient with respect to W_d is given by

$$\frac{\partial J(W)}{\partial W_d} = \frac{\partial L(W)}{\partial W_d} + \frac{\partial}{\partial W_d} \cdot \lambda \sum_{d} |W_d|$$

$$L(W) = -\sum_{i=1}^{N} y^{(i)} \cdot \log(\sigma(W \cdot x_i)) + (1 - y^{(i)}) \cdot \log(1 - \sigma(W \cdot x_i))$$

$$\frac{\partial L(w)}{\partial w_d} = -\sum_{i=1}^{N}\Big[y^{(i)} \cdot \frac{1}{\sigma(w\cdot x_i)} \cdot \sigma(w\cdot x_i)\cdot(1-\sigma(w\cdot x_i))x_{i,d}$$

$$-(1-y^{(i)}) \cdot \frac{1}{(1-\sigma(w\cdot x_i)} \cdot \sigma(w\cdot x_i)(1-\sigma(w\cdot x_i))\, x_{i,d}\Big]$$

$$= -\sum_{i=1}^{N} x_{i,d}\Big[y^{(i)} \cdot (1-\sigma(w\cdot x_i)) - (1-y^{(i)}) \cdot \sigma(w\cdot x_i)\Big]$$

$$= -\sum_{i=1}^{N} x_{i,d}\Big[y^{(i)} - y^{(i)} \cdot \sigma(w\cdot x_i) - \sigma(w\cdot x_i) + y^{(i)} \cdot \sigma(w\cdot x_i)\Big]$$

$$= \boxed{-\sum_{i=1}^{N} x_{i,d}\Big[y^{(i)} - \sigma(w\cdot x_i)\Big]} \quad \text{— } (1)$$

$$\frac{\partial}{\partial w_d} \lambda \sum_d |w_d| = \sum_d \begin{cases} -\lambda & \text{if } w_d < 0 \\ 0 & \text{if } w_d = 0 \\ \lambda & \text{if } w_d > 0 \end{cases}$$

$$= \boxed{\begin{matrix} \lambda(\text{sign}(w_d)) & \text{if } w_d \neq 0 \\ 0 & \text{if } w_d = 0 \end{matrix}} \quad (2)$$

Combining Equation 1 & 2

$$\frac{\partial J(w)}{\partial w_d} = -\sum_{i=1}^{N} x_{i,d}\left[y^{(i)} - \sigma(w \cdot x_i)\right]$$

$$+ \lambda(\text{sign}(w_d)) \quad \text{if } w_d \neq 0$$

or

$$-\sum_{i=1}^{N} x_{i,d}\left[y^{(i)} - \sigma(w \cdot x_i)\right] + 0 \quad \text{if } w_d = 0$$