- Logistic classifier (linear classifier):

$$\mathrm{WX} + \mathrm{b} = \mathrm{Y}$$

W: Weight, X: Input, B: Biased, Y: Predictions

- SOFTMAX function: turns scores into probabilities

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$$

Increase the size of output, the classifier becomes very confident about the predictions (0 or 1).

- ONE-HOT encoding: only one probability is 1, else are 0.

- CROSS ENTROPY:

$$\mathrm{D}(\mathrm{S}(\mathrm{WX} + \mathrm{b}), \mathrm{L}) = -\sum_i L_i \log(S_i)$$

S: SOFTMAX function, L: ONE-HOT encoding function D: distance

- Training loss: average of cross entropy

$$\mathrm{L} = \frac{1}{N}\sum_i D(S(WX_i + b), L_i)$$

Each distance should be small→good job of training→minimize loss function

Gradient descent: $-\alpha \nabla \mathrm{L}(weight_1, weight_2)$

$$\begin{cases} W_{i+1} = W_i - \alpha \nabla_W L = W_i - \alpha \frac{1}{N}\sum_i \frac{\partial}{\partial W} D(S(WX_i + b), L_i) \\ b_{i+1} = b_i - \alpha \nabla_b L \end{cases}$$

$$\begin{bmatrix} w_{1,1} & \dots & w_{1,400} \\ w_{2,1} & \dots & w_{2,400} \end{bmatrix} \begin{bmatrix} x_{1,1} & \cdots & x_{1,50} \\ \vdots & \ddots & \vdots \\ x_{400,1} & \cdots & x_{400,50} \end{bmatrix} + \begin{bmatrix} b_1 & \dots & b_1 \\ b_2 & \dots & b_2 \end{bmatrix} = \begin{bmatrix} y_{1,1} & \cdots & y_{1,50} \\ y_{2,1} & \cdots & y_{2,50} \end{bmatrix}$$

$\alpha\nabla_W L$

$$= \frac{\alpha}{N}\frac{\partial}{\partial W}\sum_i D(S(WX_i + b), L_i)$$

$$= -\frac{\alpha}{N}\frac{\partial}{\partial W}\sum_{50}[\sum_{2} L_i \ln(\frac{e^{y_i}}{\sum_j e^{y_j}})]$$

$$= -\frac{\alpha}{N}\frac{\partial}{\partial W}\sum_{50}\sum_{2} L_i \left[\ln\left(\frac{e^{W_iX_i+b_i}}{e^{W_iX_i+b_i} + e^{W_iX_i+b_i} + \cdots + e^{W_iX_i+b_i}}\right)\right]$$

$$= -\frac{\alpha}{N}\frac{\partial}{\partial W}\sum_{50}\sum_{2} L_{i,j}\left[\left(W_iX_j + b_i - \ln\left(e^{W_1X_j+b_1} + \cdots + e^{W_iX_j+b_i}\right)\right)\right]$$

$$= -\frac{\alpha}{N}\frac{\partial}{\partial W}\sum_{50}\sum_{2} L_{i,j}\left[\begin{bmatrix} w_{i,1} & \cdots & w_{i,400}\end{bmatrix}\begin{bmatrix} x_{1,j} \\ \vdots \\ x_{400,j}\end{bmatrix} + bi - \ln\left(e^{W_1X_j+b_1} + \cdots + e^{W_iX_j+b_i}\right)\right]$$

$$= -\frac{\alpha}{N}\frac{\partial}{\partial W}\sum_{50}\sum_{2}\left[L_{i,j}\begin{bmatrix} w_{i,1} & \cdots & w_{i,400}\end{bmatrix}\begin{bmatrix} x_{1,j} \\ \vdots \\ x_{400,j}\end{bmatrix} - L_{i,j}\ln\left(e^{W_1X_j+b_1} + \cdots + e^{W_iX_j+b_i}\right)\right]$$

$$= -\frac{\alpha}{N}\frac{\partial}{\partial W}\sum_{50}\left\{\left[L_{1,j}\begin{bmatrix} w_{1,1} & \cdots & w_{1,400}\end{bmatrix}\begin{bmatrix} x_{1,j} \\ \vdots \\ x_{400,j}\end{bmatrix} - L_{1,j}\ln\left(e^{W_1X_j+b_1} + \cdots + e^{W_iX_j+b_i}\right)\right] + \cdots + \left[L_{i,j}\begin{bmatrix} w_{i,1} & \cdots & w_{i,400}\end{bmatrix}\begin{bmatrix} x_{1,j} \\ \vdots \\ x_{400,j}\end{bmatrix} - L_{i,j}\ln\left(e^{W_1X_j+b_1} + \cdots + e^{W_iX_j+b_i}\right)\right]\right\}$$

$$= -\frac{\alpha}{N}\frac{\partial}{\partial W}\sum_{50}\left\{L_{1,j}\begin{bmatrix} w_{1,1} & \cdots & w_{1,400}\end{bmatrix}\begin{bmatrix} x_{1,j} \\ \vdots \\ x_{400,j}\end{bmatrix} + \cdots L_{i,j}\begin{bmatrix} w_{i,1} & \cdots & w_{i,400}\end{bmatrix}\begin{bmatrix} x_{1,j} \\ \vdots \\ x_{400,j}\end{bmatrix} - (L_{1,j} + \cdots + L_{i,j})\ln\left(e^{W_1X_j+b_1} + \cdots + e^{W_iX_j+b_i}\right)\right\}$$

$$= -\frac{\alpha}{N}\frac{\partial}{\partial W}\left\{\left[L_{1,1}\begin{bmatrix} w_{1,1} & \cdots & w_{1,400}\end{bmatrix}\begin{bmatrix} x_{1,1} \\ \vdots \\ x_{400,1}\end{bmatrix} + \cdots L_{i,1}\begin{bmatrix} w_{i,1} & \cdots & w_{i,400}\end{bmatrix}\begin{bmatrix} x_{1,1} \\ \vdots \\ x_{400,1}\end{bmatrix}\right] + \cdots + \left[L_{1,j}\begin{bmatrix} w_{1,1} & \cdots & w_{1,400}\end{bmatrix}\begin{bmatrix} x_{1,j} \\ \vdots \\ x_{400,j}\end{bmatrix} + \cdots L_{i,j}\begin{bmatrix} w_{i,1} & \cdots & w_{i,400}\end{bmatrix}\begin{bmatrix} x_{1,j} \\ \vdots \\ x_{400,j}\end{bmatrix}\right]\right.$$

$$\left. - \left[(L_{1,1} + \cdots + L_{i,1})\ln\left(e^{W_1X_1+b_1} + \cdots + e^{W_iX_1+b_i}\right) + \cdots + \left(L_{1,j} + \cdots + L_{i,j}\right)\ln\left(e^{W_1X_j+b_1} + \cdots + e^{W_iX_j+b_i}\right)\right]\right\}$$

$$=-\frac{\alpha}{N}\begin{bmatrix} \frac{\partial}{\partial w_{1,1}}\Psi & \cdots & \frac{\partial}{\partial w_{1,400}}\Psi \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial w_{i,1}}\Psi & \cdots & \frac{\partial}{\partial w_{i,400}}\Psi \end{bmatrix}$$

$$= -\frac{\alpha}{N}\begin{bmatrix} (L_{1,1}x_{1,1} + \cdots + L_{1,j}x_{1,j}) - [(L_{1,1} + \cdots + L_{i,1})\frac{e^{W_1X_1+b_1}}{e^{W_1X_1+b_1} + \cdots + e^{W_iX_1+b_i}}x_{1,1} + \cdots + (L_{1,j} + \cdots + L_{i,j})\frac{e^{W_1X_j+b_1}}{e^{W_1X_j+b_1} + \cdots + e^{W_iX_j+b_i}}x_{1,j}] & \cdots & \\ \vdots & \ddots & \vdots \\ (L_{i,1}x_{1,1} + \cdots + L_{i,j}x_{1,j}) - [(L_{1,1} + \cdots + L_{i,1})\frac{e^{W_iX_1+b_i}}{e^{W_1X_1+b_1} + \cdots + e^{W_iX_1+b_i}}x_{1,1} + \cdots + (L_{1,j} + \cdots + L_{i,j})\frac{e^{W_iX_j+b_i}}{e^{W_1X_j+b_1} + \cdots + e^{W_iX_j+b_i}}x_{1,j}] & \cdots & \end{bmatrix}$$

$$-\frac{\alpha}{N}\begin{bmatrix} & \cdots & (L_{1,1}x_{400,1} + \cdots + L_{1,j}x_{400,j}) - [(L_{1,1} + \cdots + L_{i,1})\frac{e^{W_1X_1+b_1}}{e^{W_1X_1+b_1} + \cdots + e^{W_iX_1+b_i}}x_{400,1} + \cdots + (L_{1,j} + \cdots + L_{i,j})\frac{e^{W_1X_j+b_1}}{e^{W_1X_j+b_1} + \cdots + e^{W_iX_j+b_i}}x_{400,j}] \\ \vdots & \ddots & \vdots \\ & \cdots & (L_{i,1}x_{400,1} + \cdots + L_{i,j}x_{400,j}) - [(L_{1,1} + \cdots + L_{i,1})\frac{e^{W_iX_1+b_i}}{e^{W_1X_1+b_1} + \cdots + e^{W_iX_1+b_i}}x_{400,1} + \cdots + (L_{1,j} + \cdots + L_{i,j})\frac{e^{W_iX_j+b_i}}{e^{W_1X_j+b_1} + \cdots + e^{W_iX_j+b_i}}x_{400,j}] \end{bmatrix}$$

$$= -\frac{\alpha}{N}\begin{bmatrix} (L_{1,1}x_{1,1} + \cdots + L_{1,j}x_{1,j}) - [S(y_{1,1})x_{1,1} + \cdots + S(y_{1,j})x_{1,j}] & \cdots & (L_{1,1}x_{400,1} + \cdots + L_{1,j}x_{400,j}) - [S(y_{1,1})x_{400,1} + \cdots + S(y_{1,j})x_{400,j}] \\ \vdots & \ddots & \vdots \\ (L_{i,1}x_{1,1} + \cdots + L_{i,j}x_{1,j}) - [S(y_{i,1})x_{1,1} + \cdots + S(y_{i,j})x_{1,j}] & \cdots & (L_{i,1}x_{400,1} + \cdots + L_{i,j}x_{400,j}) - [S(y_{i,1})x_{400,1} + \cdots + S(y_{i,j})x_{400,j}] \end{bmatrix}$$

$$= -\frac{\alpha}{N}\begin{bmatrix} (L_{1,1} - S(y_{1,1}))x_{1,1} + \cdots + (L_{1,j} - S(y_{1,j}))x_{1,j} & \cdots & (L_{1,1} - S(y_{1,1}))x_{400,1} + \cdots + (L_{1,j} - S(y_{1,j}))x_{400,j} \\ \vdots & \ddots & \vdots \\ (L_{i,1} - S(y_{i,1}))x_{1,1} + \cdots + (L_{i,j} - S(y_{i,j}))x_{1,j} & \cdots & (L_{i,1} - S(y_{i,1}))x_{400,1} + \cdots + (L_{i,j} - S(y_{i,j}))x_{400,j} \end{bmatrix}$$

$$= -\frac{\alpha}{N}\begin{bmatrix} L_{1,1} - S(y_{1,1}) & \cdots & L_{1,j} - S(y_{1,j}) \\ \vdots & \ddots & \vdots \\ L_{i,1} - S(y_{i,1}) & \cdots & L_{i,j} - S(y_{i,j}) \end{bmatrix}\begin{bmatrix} x_{1,1} & \cdots & x_{400,1} \\ \vdots & \ddots & \vdots \\ x_{1,j} & \cdots & x_{400,j} \end{bmatrix}$$

$$= -\frac{\alpha}{N}\begin{bmatrix} L_{1,1} - S(y_{1,1}) & \cdots & L_{1,j} - S(y_{1,j}) \\ \vdots & \ddots & \vdots \\ L_{i,1} - S(y_{i,1}) & \cdots & L_{i,j} - S(y_{i,j}) \end{bmatrix}X^T$$

$$\alpha\nabla_W L = \alpha\frac{\partial}{\partial w_{jk}}\sum_i D(S(WX_i+b),L_i) = -\frac{\alpha}{N}\frac{\partial}{\partial w_{jk}}\sum_i[\sum_j L_{ij}\ln(\frac{e^{y_{ij}}}{\sum_j e^{y_j}})] = -\frac{\alpha}{N}\frac{\partial}{\partial w_{jk}}\sum_i[\sum_j L_{ij}\,y_{ij} - L_{ij}\ln(\sum_j e^{y_j})]$$

$$= -\frac{\alpha}{N}\frac{\partial}{\partial w_{jk}}\sum_i[\sum_j L_{ij}\,(\sum_k\left(w_{jk}x_{ki}\right)+b_j) - L_{ij}\ln(e^{y_{1i}}+\cdots+e^{y_{ji}})]$$

$$= -\frac{\alpha}{N}\sum_i[L_{ij}x_{ki} - \sum_j L_{ij}(\mathrm{S}(y_{1i})+..+\mathrm{S}(y_{ji}))] = -\frac{\alpha}{N}\sum_i(L_{ij}-\mathrm{S}_{ij})x_{ki}$$

$\alpha\nabla_B L$

$$= -\frac{\alpha}{N}\frac{\partial}{\partial B}\sum_{50}\sum_{2}L_{i,j}\left[\begin{bmatrix}w_{i,1} & \cdots & w_{i,400}\end{bmatrix}\begin{bmatrix}x_{1,j}\\ \vdots \\ x_{400,j}\end{bmatrix} + bi - \ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)\right]$$

$$= -\frac{\alpha}{N}\begin{bmatrix}\frac{\partial}{\partial b_1}\sum_{50}\sum_2 L_{i,j}\left[\begin{bmatrix}w_{i,1} & \cdots & w_{i,400}\end{bmatrix}\begin{bmatrix}x_{1,j}\\ \vdots \\ x_{400,j}\end{bmatrix} + bi - \ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)\right] \\ \vdots \\ \frac{\partial}{\partial b_i}\sum_{50}\sum_2 L_{i,j}\left[\begin{bmatrix}w_{i,1} & \cdots & w_{i,400}\end{bmatrix}\begin{bmatrix}x_{1,j}\\ \vdots \\ x_{400,j}\end{bmatrix} + bi - \ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)\right]\end{bmatrix}$$

$$= -\frac{\alpha}{N}\begin{bmatrix}\frac{\partial}{\partial b_1}\sum_{50}\sum_2 L_{i,j}\left[+bi - \ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)\right] \\ \vdots \\ \frac{\partial}{\partial b_i}\sum_{50}\sum_2 L_{i,j}\left[+bi - \ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)\right]\end{bmatrix} = -\frac{\alpha}{N}\begin{bmatrix}\frac{\partial}{\partial b_1}\sum_{50}\sum_2\left[L_{i,j}bi - L_{i,j}\ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)\right] \\ \vdots \\ \frac{\partial}{\partial b_i}\sum_{50}\sum_2\left[L_{i,j}\,bi - L_{i,j}\ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)\right]\end{bmatrix}$$

$$= -\frac{\alpha}{N}\begin{bmatrix}\frac{\partial}{\partial b_1}\sum_{50}\{[L_{1,j}b1 - L_{1,j}\ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)]+\cdots+[L_{i,j}bi - L_{i,j}\ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)]\} \\ \vdots \\ \frac{\partial}{\partial b_i}\sum_{50}\{[L_{1,j}b1 - L_{1,j}\ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)]+\cdots+[L_{i,j}bi - L_{i,j}\ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)]\}\end{bmatrix}$$

$$= -\frac{\alpha}{N}\begin{bmatrix}\frac{\partial}{\partial b_1}\sum_{50}\{[L_{1,j}b1+\cdots+L_{i,j}bi] - (L_{1,j}+\cdots+L_{i,j})\ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)\} \\ \vdots \\ \frac{\partial}{\partial b_i}\sum_{50}\{[L_{1,j}b1+\cdots+L_{i,j}bi] - (L_{1,j}+\cdots+L_{i,j})\ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)\}\end{bmatrix}$$

$$= -\frac{\alpha}{N}\begin{bmatrix}\frac{\partial}{\partial b_1}\{\left[L_{1,1}b1+\cdots+L_{i,1}bi\right]+\cdots+[L_{1,j}b1+\cdots+L_{i,j}bi] - (L_{1,1}+\cdots+L_{i,1})\ln(e^{W_1X_1+b_1}+\cdots+e^{W_iX_1+b_i}) - \cdots - (L_{1,j}+\cdots+L_{i,j})\ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)\} \\ \vdots \\ \frac{\partial}{\partial b_i}\{\left[L_{1,1}b1+\cdots+L_{i,1}bi\right]+\cdots+[L_{1,j}b1+\cdots+L_{i,j}bi] - (L_{1,1}+\cdots+L_{i,1})\ln(e^{W_1X_1+b_1}+\cdots+e^{W_iX_1+b_i}) - \cdots - (L_{1,j}+\cdots+L_{i,j})\ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)\}\end{bmatrix}$$

$$= -\frac{\alpha}{N}\begin{bmatrix}\left[L_{1,1}+\cdots+L_{1,j}\right] - \{\left(L_{1,1}+\cdots+L_{i,1}\right)\frac{\partial}{\partial b_1}\ln(e^{W_1X_1+b_1}+\cdots+e^{W_iX_1+b_i})+\cdots+(L_{1,j}+\cdots+L_{i,j})\frac{\partial}{\partial b_1}\ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)\} \\ \vdots \\ \left[L_{i,1}+\cdots+L_{i,j}\right] - \{\left(L_{1,1}+\cdots+L_{i,1}\right)\frac{\partial}{\partial b_i}\ln(e^{W_1X_1+b_1}+\cdots+e^{W_iX_1+b_i})+\cdots+(L_{1,j}+\cdots+L_{i,j})\frac{\partial}{\partial b_i}\ln\left(e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}\right)\}\end{bmatrix}$$

$$= -\frac{\alpha}{N}\begin{bmatrix}\left[L_{1,1}+\cdots+L_{1,j}\right] - \left\{\left(L_{1,1}+\cdots+L_{i,1}\right)\frac{e^{W_1X_1+b_1}}{e^{W_1X_1+b_1}+\cdots+e^{W_iX_1+b_i}}+\cdots+\left(L_{1,j}+\cdots+L_{i,j}\right)\frac{e^{W_1X_j+b_1}}{e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}}\right\} \\ \vdots \\ \left[L_{i,1}+\cdots+L_{i,j}\right] - \left\{\left(L_{1,1}+\cdots+L_{i,1}\right)\frac{e^{W_iX_1+b_i}}{e^{W_1X_1+b_1}+\cdots+e^{W_iX_1+b_i}}+\cdots+\left(L_{1,j}+\cdots+L_{i,j}\right)\frac{e^{W_iX_j+b_i}}{e^{W_1X_j+b_1}+\cdots+e^{W_iX_j+b_i}}\right\}\end{bmatrix}$$

$$= -\frac{\alpha}{N}\begin{bmatrix}\left[L_{1,1}+\cdots+L_{1,j}\right] - \{\left(L_{1,1}+\cdots+L_{i,1}\right)S\left(y_{1,1}\right)+\cdots+\left(L_{1,j}+\cdots+L_{i,j}\right)S\left(y_{1,j}\right)\} \\ \vdots \\ \left[L_{i,1}+\cdots+L_{i,j}\right] - \{\left(L_{1,1}+\cdots+L_{i,1}\right)S\left(y_{i,1}\right)+\cdots+\left(L_{1,j}+\cdots+L_{i,j}\right)S\left(y_{i,j}\right)\}\end{bmatrix} = -\frac{\alpha}{N}\begin{bmatrix}\left[L_{1,1}+\cdots+L_{1,j}\right] - \left\{S\left(y_{1,1}\right)+\cdots+S\left(y_{1,j}\right)\right\} \\ \vdots \\ \left[L_{i,1}+\cdots+L_{i,j}\right] - \left\{S\left(y_{i,1}\right)+\cdots+S\left(y_{i,j}\right)\right\}\end{bmatrix}$$

$$\alpha\nabla_B L = \alpha\frac{\partial}{\partial b_j}\sum_i D(S(WX_i+b),L_i) = -\frac{\alpha}{N}\frac{\partial}{\partial b_j}\sum_i[\sum_j L_{ij}\ln(\frac{e^{y_{ij}}}{\sum_j e^{y_j}})] = -\frac{\alpha}{N}\frac{\partial}{\partial b_j}\sum_i[\sum_j L_{ij}\,y_{ij} - L_{ij}\ln(\sum_j e^{y_j})]$$

$$= -\frac{\alpha}{N}\frac{\partial}{\partial b_j}\sum_i[\sum_j L_{ij}\,(\sum_k\left(w_{jk}x_{ki}\right)+b_j) - L_{ij}\ln(e^{y_{1i}}+\cdots+e^{y_{ji}})]$$

$$= -\frac{\alpha}{N}\sum_i[L_{ij} - \sum_j L_{ij}(\mathrm{S}(y_{1i})+..+\mathrm{S}(y_{ji}))] = -\frac{\alpha}{N}\sum_i L_{ij} - \mathrm{S}_{ij}$$
