Multiclass logistic Regression - with regularizatn

$$P(C_k/\phi) = y_k(\phi) = \frac{\exp(a_k)}{\sum_j \exp(a_j)}$$

$$a_k = w_k^T \phi$$

$$\frac{\partial y_k}{\partial a_j} = y_k(I_{kj} - y_j)$$

$$P(T/w_1, w_2 \ldots w_k) = \prod_{n=1}^{N} \prod_{\cancel{k}=1}^{K} P(C_k/\phi_n)^{t_{nk}}$$

$$= \prod_{n=1}^{N} \prod_{k=1}^{K} y_{nk}^{t_{nk}}$$

Log likelihood :-

$$\log P(T/w_1, w_2 \ldots w_k) = \sum_{n=1}^{N} \sum_{\cancel{n}k=1}^{K} t_{nk} \cancel{\ln} \ln y_{nk}$$

Cross entropy (-ve of log likelihood)

for L_2 regularized one

Regularized log likelihood

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \ln y_{nk} - \cancel{\lambda \|w\|} \frac{\lambda}{2} \sum_{i=1}^{n} w_i^2$$

Scanned with CamScanner

1st derivative

$$\frac{t_{nk} \ln y_{nk}}{\partial w_j^{\circ}} = \frac{\partial t_{nk} \ln y_{nk}}{\partial y_{nk}} \frac{\partial y_{nk}}{\partial a_j} \cdot \frac{\partial a_j}{\partial w_j^{\circ}}$$

$$= t_{nk} \frac{1}{y_{nk}} y_{nk}(I_{kj} - y_{nj}) \phi_n$$

$$= t_{nk}(I_{kj} - y_{nj}) \phi_n$$

$$+\sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} (I_{kj} - y_{nj}) \phi_n$$

$$= -\sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} y_n \phi_n + \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} I_{kj} \phi_n$$

$$= \sum_{n=1}^{N} \left(\left(-\left(\sum_{k=1}^{K} t_{nk}\right) y_{nj} \phi_n + \sum_{n=1}^{N} t_{nj} \phi_n \right) \right.$$

$$= + \sum_{n=1}^{N} \phi_n (t_{nj} - y_{nj})$$

With regularizatn.

$$\sum_{n=1}^{N} \phi_n (t_{nj} - y_{nj}) - \lambda \sum_{i=1}^{N} w_i^{\circ}$$

2nd derivative

$$+ \sum_{n=1}^{N} y_{nk} (I_{kj} - y_{nj}) \phi_n \phi_n^{T} - \lambda$$