3) In the case of softmax, we have a weight vector corresponding to the individual class.

$$p(y_n = k | x_n, w) = \frac{\exp(w_k^T x_n)}{\sum_{l=1}^{K} \exp(w_l^T x_n)} = \mu_{nk}$$

In this we have the likelihood function as

$$p(y|x, w) = \prod_{n=1}^{N} \prod_{l=1}^{k} \mu_{nl}^{y_{nl}}$$

Now taking log we get

$$\log(p(y|x, w)) = \log \prod_{n=1}^{N} \prod_{l=1}^{k} \mu_{nl}^{y_{nl}}$$

$$= \sum_{n=1}^{N} \sum_{l=1}^{k} \log \mu_{nl}^{y_{nl}}$$

$$= \sum_{n=1}^{N} \sum_{l=1}^{k} \left[y_{nl} w_l^T x_n - y_{nl} \log \sum_{m=1}^{k} e^{w_m^T x_n} \right]$$

$$f(w) = LL = \sum_{n=1}^{N} \left[\sum_{l=1}^{k} y_{nl} w_l^T x_n - y_{nl} k \log \sum_{m=1}^{k} e^{w_m^T x_n} \right]$$

→ Now taking the derivative w.r.t w_k.

$$\frac{\partial f}{\partial w_i} = \sum_{n=1}^{N} \left[y_{ni} x_n - \frac{k y_{ni} x_n e^{(w_i^T x_n)}}{\sum_{m=1}^{k} \exp(w_m^T x_n)} \right]$$

→ So the above is the partial derivative

Scanned by CamScanner

So now doing gradient decent we have

$$w_i = w_i - \frac{\partial F}{\partial w_i} \cdot \eta_i$$

$$\Rightarrow w_i = w_i - \sum_{n=1}^{N} \left(y_{n_i} x_n - \frac{x_n \cdot e^{w_i^T x_n}}{\sum_{m=1}^{k} \exp(w_m^T x_n)} \right)$$

Intuitive meaning.

$\rightarrow$ $y_{nk} - \mu_{nk}$ depicts errors

so if $y_{nk} = 1$ and prob. μ_{nk} is low

then $y_{nk} - \mu_{nk}$ will give more penality more.

- for error, there would be high penalizatio
- for correct prediction little penalizatio

4) Path 1

Assume that the two sets of points are linearly separable, this means that we have vector w_0 and a constant a such that

for all x_n in $\{x_1, x_2 \dots x_N\}$

$$w_0^T x_n + a > 0$$

for all y_n in $\{y_1, y_2 \dots y_M\}$

$$w_0^T y_n + a < 0$$

Now the convex hull of $\{x_1, x_2 \dots x_N\}$ is a structure that has all the points of the form $\sum_{n=1}^{N} \alpha_n x_n$ and $\sum_{n=1}^{N} \alpha_n = 1$, $\alpha_n \geq 0$

Now consider any point x present in the convex hull of $\{x_1, x_2 \dots x_N\}$

$$x = \sum_{n=1}^{N} \alpha_n x_n \quad \sum_{n=1}^{N} \alpha_n = 1 \quad \alpha_n \geq 0$$

Consider

$$w_0^T x + a$$

$$= w_0^T \sum \alpha_n x_n + a$$

$$= \sum \alpha_n (w_0^T x_n) + a$$

$$= \sum \alpha_n [w_0^T x_n + a] \qquad \left\{ \sum_{n=1}^{N} \alpha_n a = a \right.$$

$$\therefore w_0^T x + a > 0$$

Now consider the convex hull of $\{y_1, y_2 \cdots y_M\}$

Every point y in the convex hull is of the form

$$y = \sum_{n=1}^{M} B_n y_n \quad , B_n \geq 0 \; , \sum_{n=1}^{M} B_n = 1$$

Consider

$$w_0^T y + a$$

$$= w_0^T \sum_{n=1}^{M} B_n y_n + a$$

$$= \sum_{n=1}^{M} B_n [w_0^T y_n + a]$$

$$\Rightarrow \quad w_0^T y + a < 0 \quad \text{for all } y.$$

hence we have shown that all points of convex hull of $\{x_1, x_2 \cdots x_N\}$ lie on one side and all point in convex hull of $\{y_1, y_2 \cdots y_M\}$ lie on another side.

Hence the convex hull do not intersect.

→ Assume that the convex hulls do not intersect.

Now since the two convex hull do not intersect, we have a min. distance d between the two convex hulls

So since the min distance is d, $\exists$ point $p_1 \in C_1$, $p_2 \in C_2$ such that

$$|p_1 - p_2| = d.$$

Since convex hulls don't intersect, then solving SVM dual is same as finding a linear separator which is same as finding the shortest line joining convex hull since the separator is the $\perp$ bisector.