Machine Learning HW#4

Evgeny Marshakov

Problem 1

Let X, Y be a convex hulls of the sets $\{x_n\}$ and $\{y_n\}$, resp., i.e

$$X = \left\{ \sum_n a_n x_n : \sum_n a_n = 1, a_n \geq 0 \right\} \tag{1}$$

$$Y = \left\{ \sum_m b_m y_m : \sum_m b_m = 1, b_m \geq 0 \right\} \tag{2}$$

If these convex hulls intersect then there exists a point s.t.

$$z = \sum_n a_n x_n = \sum_m b_m y_m \tag{3}$$

$$a_n, b_m \geq 0, \ \sum_n a_n = \sum_m b_m = 1 \tag{4}$$

Let us assume that the sets $\{x_n\}$ and $\{y_n\}$ are linearly separable, i.e. there are $\hat{w}$ and w_0. s.t.

$$\hat{w} \cdot x_n + w_0 > 0, \ \forall n \tag{5}$$

$$\hat{w} \cdot y_m + w_0 < 0, \ \forall m \tag{6}$$

Hence we obtain the following

$$\hat{w} \cdot z + w_0 = \sum_n a_n \hat{z} \cdot x_n + w_0 = \sum_n a_n (\hat{w} \cdot x_n + w_0) > 0 \tag{7}$$

$$\hat{w} \cdot z + w_0 = \sum_m b_m \hat{z} \cdot y_m + w_0 = \sum_m b_m (\hat{w} \cdot y_m + w_0) < 0 \tag{8}$$

So we obtain a contradiction and hence $\{x_n\}$ and $\{y_m\}$ cannot be linearly separable. Conversely, if these sets are linearly separable, then we can't find any point in the intersection of their convex hulls due to inequalities (7) and (8) (also we know that A implies B iff $\overline{B}$ implies $\overline{A}$).

Problem 2

Since our data is linearly separable then any decision boundary will have the following property

$$w^T\phi_n = \begin{cases} \geq 0, & t_n = 1 \\ < 0, & \text{otherwise} \end{cases} \tag{9}$$

Also it can be easily seen that the negative log-likelihood function

$$E(w) = -\sum_{n=1}^{N} \left(t_n \ln y_n + (1-t_n)\ln(1-y_n)\right) \tag{10}$$

is minimized when $y_n = \sigma(w^T\phi_n) = t_n$. This result can be obtained when the sigmoid function is very close to the Heaviside step function, more precisely, when the argument tends to $\pm\infty$ or when the magnitude of w tends to ∞.

Problem 3

We know that

$$\nabla E = \sum_n (y_n - t_n)\phi_n \tag{11}$$

$$\mathbb{H} = \nabla\nabla E = \sum_n y_n(1-y_n)\phi_n\phi_n^T \tag{12}$$

It can be easily seen that if $u \neq 0$ then

$$u^T\mathbf{H}u = u^T\Phi^T R\Phi u = u^T\Phi^T R^{\frac{1}{2}} R^{\frac{1}{2}}\Phi u = \|R^{\frac{1}{2}}\Phi u\|^2 > 0 \tag{13}$$

So the Hessian matrix is positive definite (we know that Φ is non degenerate). Hence, the error function $E(w)$ is strictly convex, so it has only one minimum.