Parsa Daghigh 81010419 ML - HW1

① A) Why cross-validation?

We use cross-validation to track the models performance and generalizability. Helps to reduce overfitting. And insure the model is robust and reliable for new data.

Methods : I) K-Fold dataset $\xrightarrow{\text{divided to}}$ K-folds (parts)

model is trained on k-1 folds and the last fold as test

Each time a fold is used for test. (k-times this process happens)

Average of these k-acc is the result. And reducing variance is important.

II) Stratified K-Fold

very ideal for unbalanced datasets. Similar to K-fold but the difference is that each fold holds the representation of all datas. For example if 70% are 'Cat' Labled and 30% 'dog' Labled. then 70% in every fold should be Cat and 30% dog.

III) Leave-out-one Cross-validation

Each data point is considered once as a test. After all process, the average of all process is calculated. It is computationaly hard.

B) $D(x,y) = \sqrt{\sum_{k=1}^{d}(x_k - y_k)^2}$ if $x'_k = a_k x_k$ for $k=1,2,\dots,d$

Is it still a Eucilidian metric?

$$D(x',y) = \sqrt{\sum_{k=1}^{d}(\alpha_k x_k - \alpha_k y_k)^2} = \sqrt{\sum_{k=1}^{d}\alpha_k^2(x_k - y_k)^2}$$

Standard metric :

1) $D(x,x) = 0 \iff \sqrt{\sum_{k=1}^{d}\alpha_k^2(x_k - x_k)^2} = \sqrt{\sum_{k=1}^{d}\alpha_k^2(0)^2} = 0$

2) $D(x,y) = D(y,x) \iff \sqrt{\sum_{k=1}^{d}\alpha_k^2(x_k - y_k)^2} = \sqrt{\sum_{k=1}^{d}\alpha_k^2(y_k - x_k)^2}$

3) if $x \neq y$ $D(x,y) > 0 \iff \sqrt{\sum_{k=1}^{d}\alpha_k^2(x_k - y_k)^2} > 0$

4) $D(x,y) + D(y,z) \geq D(x,z)$

$$\iff \sqrt{\sum_{k=1}^{d}\alpha_k^2(x_k - y_k)^2} + \sqrt{\sum_{k=1}^{d}\alpha_k^2(y_k - z_k)^2} > \sqrt{\sum_{k=1}^{d}\alpha_k^2(x_k - z_k)^2}$$

توان ۲

$$\iff \sum_{k=1}^{d}\alpha_k^2(x_k - y_k)^2 + \sum_{k=1}^{d}\alpha_k^2(y_k - z_k)^2 + 2\sqrt{\sum_{k=1}^{d}\alpha_k^2(x_k - y_k)^2}\cdot\sqrt{\sum_{k=1}^{d}\alpha_k^2(y_k - z_k)^2} \geq \sum_{k=1}^{d}\alpha_k^2(x_k - z_k)^2$$

Cauchy-Schwarz : $\left(\sum_{k=1}^{d}\alpha_k^2(x_k - y_k)^2\right)\cdot\left(\sum_{k=1}^{d}\alpha_k^2(y_k - z_k)^2\right) \geq \left(\sum_{k=1}^{d}\alpha_k^2(x_k - y_k)(y_k - z_k)\right)^2 = *$

با نامساوی بالا

$$\iff \sum_{k=1}^{d}\alpha_k^2\left((x_k - y_k)^2 + (y_k - z_k)^2\right) + 2\underbrace{\sum_{k=1}^{d}\alpha_k^2(x_k - y_k)(y_k - z_k)}_{\sqrt{*}} = \sum_{k=1}^{d}\alpha_k^2(x_k - y_k + y_k - z_k)^2$$

$$= \sum_{k=1}^{d}\alpha_k^2(x_k - z_k)^2$$

$\iff D(x,y) + D(y,z) \geq D(x,z)$