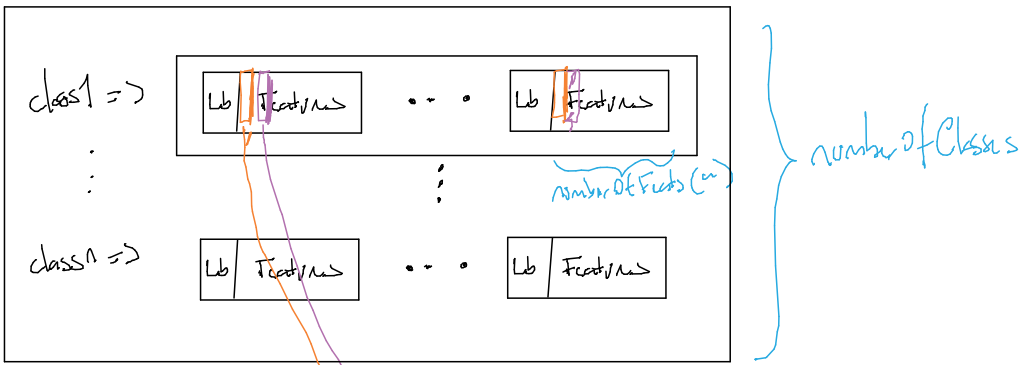
ReliefF

nearestNeighborsPerClass : Map[Int, Array[LabeledPoint]]

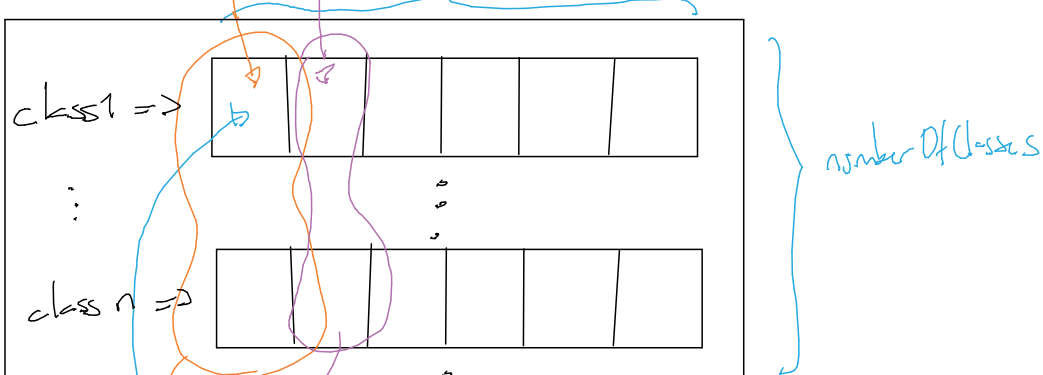


numberOfNeighbors (k)

sum
diff

sumOfDiffsPerClass : Map[Int, IndexedSeq[Double]]

numberOfFeats (a)



sumOfDiffs (A, Ri, Ni)

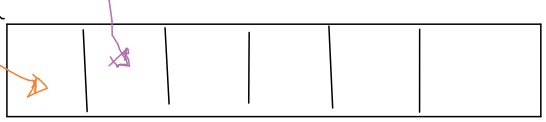
Singles (m)

1 ... numberOfNeighbors

weightedSum

tWeights: Array[Double]

numberOfFeats (a)



$$X = \{ (\vec{F}_1, c_1), (\vec{F}_2, c_2), \dots (\vec{F}_n, c_n) \}$$

$$R = (R_1, R_2, \dots, R_m)$$

$R_i = (\vec{F}_i, c_i)$

$X \Rightarrow$ map $\Rightarrow$ XD $\Rightarrow$ takeOrdered $\Rightarrow$ KNN

$$MAX = \{ \max(F_{1,1}[1] \dots F_{n,1}[1]), \dots, \max(F_{1,q}[q] \dots F_{n,q}[q]) \}$$

$$MAX_i = \max(F_{1,i}, \dots F_{n,i})$$ ✓

$$XD = \{ ((\vec{F}_1, c_1), \vec{D}_1), \dots ((\vec{F}_n, c_n), \vec{D}_n) \}$$

$$D_i = (\text{distance}(F_i, R_1), \dots, \text{distance}(F_i, R_m))$$

$$NNC_i = \{ (c_1, \vec{NN}_{i,1}), \dots (c_c, \vec{NN}_{i,c}) \}$$

↓ nearest Neighbors of the i-esima sample for the 1-esima class.

$$SDC_i = \{ (c_1, \vec{SD}_{i,1}), \dots (c_c, \vec{SD}_{i,c}) \}$$

$|SD_{i,j}| = a$

$$SD_{i,j,A} = \sum_{Y=1}^{K} diff(\vec{NN}_{i,j,A}, R_{i,1,A})$$

$1 \dots q$

facts of i-esima sample — A-esima fact

$$W_A = \sum_{i=1}^{m} \left[\sum_{\substack{z=1, \\ z \neq R_{Y,2}}}^{c} \frac{p(z)}{1 - p(R_{i,2})} \cdot SD_{Y,z,A} - SD_{i,R_{i,2},A} \right]$$

$p(R_{i,2})$: class

Σ

dataWithDistances: RDD[LabeledPoint, Array[Double]]