

CS6140 Machine Learning Fall 2014 Homework 1, Wei Luo

PROBLEM 3

Suppose we have an arbitrary decision tree T . And there are two nodes n_1, n_2 that have same splits (*feature f , threshold t*) on a root-leaf path of that tree. And n_2 is in one subtree of n_1 .

Since n_2 is in a subtree of n_1 , the data points have already been splitted by n_1 , they must have either $f < t$ or $f \geq t$. So for all the data points at n_2 , the decision should all go to one side of the sub trees.

Now we change the tree by removing the node n_2 and its subtree that no data points can reach, and put the root node of the other subtree at the position of n_2 . By doing this we get a new decision tree T' which is equivalent to T . Because for any data point, they will get same decision result (prediction) since we are just removing decision branches that could never be reached.

For all repeated splits in an arbitrary tree, we can repeat the actions above to get a new decision tree only with distinct splits on each path. And the new decision tree is equivalent to the original tree as described above.

a) For an arbitrary decision tree T with unequal branching ratios. For any node that have multiple sub-branches we can first make two groups of the sub-branches and do a binary split at this node. Then at next level, we split the groups we get before. Repeat these steps until we get all the branches at the beginning classified in a binary sub-tree. Then we keep all successors as before. Do this for every node that have multiple sub-branches, we can get an equivalent binary decision tree to T .

b) For an equivalent binary tree with B leaf nodes, the number of levels is maximized when the tree is shaped almost linear and it is minimized when the tree is balanced and almost full.

For the almost linear binary tree with B leaf nodes, it has B levels. For an almost full balanced binary tree with n levels, it will have at most 2^n leaf nodes, so $B \leq 2^n$. We can get $n > \lfloor \log_2 B \rfloor + 1$.

So the upper limit of levels is B and lower level of limits is $\lfloor \log_2 B \rfloor + 1$.

c) A functional equivalent binary tree will have exactly $2B - 1$ nodes regardless of the tree structure. Since it will always have B leaf nodes. We can prove this by induction. When B is 2, it is obvious that the tree will have 3 nodes ($2B - 1$). Suppose when $B = k$ the tree will have $2k - 1$ nodes. When $B = k + 1$, we just want to add one leaf node to the tree. To achieve this, we have to add two leaf nodes to be successors of one of the original leaf nodes. By doing this, we added two nodes to the tree. Then the number of tree nodes becomes $2k - 1 + 2 = 2(k + 1) - 1$. The hypothesis holds when $B = k + 1$. We can conclude that for any binary tree with B leaf nodes, it will have $2B - 1$ nodes.

So the upper and lower limits of nodes of the tree are both $2B - 1$.

PROBLEM 4

a) For a binary yes/no feature, before the split, the entropy is maximum when yes/no happens equally likely. It is $\frac{1}{2} \log_2 2 + \frac{1}{2} \log_2 2 = 1$

After the split, the entropy can reach the minimum 0 when every branch is pure. Since when calculating the entropy for the branches, a $\log_2 1$ is multiplied for every branch.

So the decrease in entropy by a split on a binary feature can be at most 1 with the circumstances described above.

b) For arbitrary branching B , the entropy is maximum when everything happens equally likely. It is $\sum_1^B \frac{1}{B} \log_2 B = \log_2 B$

After the split, it is the same, the entropy can reach the minimum 0 when every branch is pure.

So the decrease in entropy by a split on an arbitrary branching B can be at most $\log_2 B$

PROBLEM 5

$$\begin{aligned}
 \begin{pmatrix} a \\ b \end{pmatrix} &= w = \left(\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix}^T \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix}^T \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \\
 &= \begin{pmatrix} n & \sum_{i=1}^m x_i \\ \sum_{i=1}^m x_i & \sum_{i=1}^m x_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m x_i y_i \end{pmatrix} \\
 &= \frac{1}{n \sum_{i=1}^m x_i^2 - (\sum_{i=1}^m x_i)^2} \begin{pmatrix} \sum_{i=1}^m x_i^2 & -\sum_{i=1}^m x_i \\ -\sum_{i=1}^m x_i & n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m x_i y_i \end{pmatrix} \\
 &= \frac{1}{n \sum_{i=1}^m x_i^2 - (\sum_{i=1}^m x_i)^2} \begin{pmatrix} \sum_{i=1}^m x_i^2 \sum_{i=1}^m y_i - \sum_{i=1}^m x_i \sum_{i=1}^m x_i y_i \\ n \sum_{i=1}^m x_i y_i - \sum_{i=1}^m x_i \sum_{i=1}^m y_i \end{pmatrix}
 \end{aligned}$$

So,

$$\begin{aligned}
 a &= \frac{\sum_{i=1}^m x_i^2 \sum_{i=1}^m y_i - \sum_{i=1}^m x_i \sum_{i=1}^m x_i y_i}{n \sum_{i=1}^m x_i^2 - (\sum_{i=1}^m x_i)^2} \\
 b &= \frac{n \sum_{i=1}^m x_i y_i - \sum_{i=1}^m x_i \sum_{i=1}^m y_i}{n \sum_{i=1}^m x_i^2 - (\sum_{i=1}^m x_i)^2}
 \end{aligned}$$

PROBLEM 7

For two sets of vectors \mathbf{x} and \mathbf{y} , their convex hulls are: $\mathbf{x} = \sum_i \alpha_i x_i$, $\mathbf{y} = \sum_j \beta_j y_j$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $\sum_i \alpha_i = 1$, $\sum_j \beta_j = 1$

First, we suppose their convex hulls intersect, there must exist a \mathbf{z} so that:

$$\mathbf{z} = \sum_i \alpha_i x_i = \sum_j \beta_j y_j$$

If the two vectors are linearly separable, there must exist a hyper plane \mathbf{w}, w_0 so that:

$$\forall i \ w^t x_i + w_0 > 0, \forall j \ w^t y_j + w_0 < 0$$

Then we calculate $w^t z + w_0$, since $\sum_i \alpha_i = 1$, $\sum_j \beta_j = 1$

$$w^t z + w_0 = w^t \sum_i \alpha_i x_i + w_0 = w^t \sum_i \alpha_i x_i + w_0 \sum_i \alpha_i = \sum_i \alpha_i (w^t x_i + w_0)$$

$$w^t z + w_0 = w^t \sum_j \beta_j y_j + w_0 = w^t \sum_j \beta_j y_j + w_0 \sum_j \beta_j = \sum_j \beta_j (w^t y_j + w_0)$$

$$\text{Then } \sum_i \alpha_i (w^t x_i + w_0) = \sum_j \beta_j (w^t y_j + w_0)$$

For $\forall i$ and $\forall j$, $w^t x_i + w_0 > 0$, $w^t y_j + w_0 < 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$

Together with the equation above, we get $\alpha_i = 0 \ \forall i$, $\beta_j = 0 \ \forall j$

But $\sum_i \alpha_i = 1$, $\sum_j \beta_j = 1$, this cannot be true with $\alpha_i = 0$ and $\beta_j = 0$.

We get a contradiction here, so if the convex hulls of two vectors intersect the two vectors are not linearly separable.

On the other hand:

Suppose the two vectors are linearly separable, there must exist a hyper plane \mathbf{w}, w_0 so that:

$$\forall i \ w^t x_i + w_0 > 0, \forall j \ w^t y_j + w_0 < 0$$

If their convex hulls intersect, the intersection \mathbf{z} satisfies

$$\mathbf{z} = \sum_i \alpha_i x_i = \sum_j \beta_j y_j$$

Then we can also get:

$$w^t z + w_0 = \sum_i \alpha_i (w^t x_i + w_0), \ w^t z + w_0 = \sum_j \beta_j (w^t y_j + w_0)$$

Since $\alpha_i \geq 0$, $\beta_j \geq 0$, $\sum_i \alpha_i = 1$, $\sum_j \beta_j = 1$, $w^t x_i + w_0 > 0$, $w^t y_j + w_0 < 0$

$$\text{Then } w^t z + w_0 = \sum_i \alpha_i (w^t x_i + w_0) > 0, \ w^t z + w_0 = \sum_j \beta_j (w^t y_j + w_0) < 0$$

Which is impossible, so if the two vectors are linearly separable, their convex hulls cannot intersect .

Thus we proved for two sets of vectors, either they are linearly separable or their convex hulls intersect.