HW1

Sicheng Wang

January 22, 2019

1 PROBLEM 3

1.1 DHS chapter8, Pb1.

Suppose a decision tree has a path with repeated splits s_1 and s_2, where $s_1 = s_2 = (f, t)$.

Say that s_1 is prior to s_2 (s_1 is closer to the root).

Suppose at split s_1, this path goes to the left branch, which means all the data subject to this path have feature f less than threshold t.

Then, at split s_2, these data will again all fall into the left path. So s_2 has an empty right branch. Therefore, removing s_2 from this path will have no impact on the decision result.

Similarly, if this path goes to the right branch at s_1, removing s_2 from this path will have no impact on the decision result.

Consequently, an equivalent decision tree only with distinct splits on each path can be constructed by iteratively removing all the repeated splits which is closer to the leaves.

1.2 DHS chapter8, 2.(a)

For a B-way split, if this feature is numeric, then this split can be treated as cutting the feature value range into B sections. In this case, it can be represented with $B-1$ binary splits, from low to high. If the feature is categorical, this split can be represented by $B-1$ 'equals' split. Therefore a B-way split, the same function can be achieved by replacing it with $B-1$ binary splits, which will also produce B possible leaves, each of which corresponding to one of the childe nodes of the B-way split.

That is to say,given an arbitraty decision tree, all of its binary non-binary split nodes can be replaced with binary splits. By doing so, this new decision tree is a binary tree because it contains only binary splits.

1.3 DHS chapter8, 2.(b)

For a B-way split, there are B possible outcomes, since one binary split will only introduce one more outcome, there needs to be at least $B-1$ nodes. which will take at least $log_2(B-1)$ levels (by splitting a numerical feature space in a binary search fashion). As for the upper bound, the can be as many as $B-1$ splits in a row, in a root-leaf path (no repeating split), which can take as much as $B-1$ levels.

1.4 DHS chapter8, 2.(c)

If the split is limited to a single feature: As demonstrated in b), the lower bound for number of nodes is $B-1$. The upper bound case is $B-1$, when all thresholds are encountered from low to high in a path (all the other branches are leaf nodes). If the split is not limited to a single feature: the lower bound is $B-1$. And the upper bount is 2^B-1, in which case the deicision tree is a complete tree.

2 PROBLEM 4

2.1 (a)

$$\begin{aligned}
\Delta i(N) =& i(N) - P_L i(N_L) - (1-P_L) i(N_R) \\
=& -\Sigma_j[\frac{N_j}{N} log_2(\frac{N_j}{N})] + \frac{N_L}{N}\Sigma_j[\frac{N_{jL}}{N_L} log_2(\frac{N_{jL}}{N_L})] + \frac{N-N_L}{N}\Sigma_j[\frac{N_j - N_{jL}}{N-N_L} log_2(\frac{N_j - N_{jL}}{N-N_L})] \\
=& -\frac{1}{N}\Sigma_j\left([N_j(log_2 N_j - log_2 N) - N_{jL}(log_2 N_{jL} - log_2 N_L) - (N_j - N_{jL})(log_2(N_j - N_{jL}) - log_2(N-N_L))]\right) \\
=& -\frac{1}{N}\Sigma_j\left([N_j log_2 N_j - N_{jL} log_2 N_{jL} - (N_j - N_{jL}) log_2(N_j - N_{jL})]\right) \\
&+ \frac{1}{N}\Sigma_j\left([N_j log_2 N - N_{jL} log_2 N_L - (N_j - N_{jL}) log_2(N - N_L)]\right)
\end{aligned}$$

The first part is a convex function, which is maximized to 0 when $N_{jL} = 0$ or $N_{jL} = N_j$, depending on N_L because $\Sigma_i N_{jL} = N_L$.

The second part is a convex function, which is mimimized to 1 when $N_L = \frac{N}{2}$.

Finally, this equation is maximized to 1 when the given data is seperated to two subsets with equal size, and there are no labels appear in both subsets.

In another way, "yes/no" query means a binary split, the data points decoded by this point either go to the left branch or the right branch. This infomation takes up to 1 bit to encode(maximum of 1 if they are splited equally and one label never goes to two braches, otherwise it's less informative).

2.2 (b)

The decrease in entropy impurity provided by a single B-way split can never be greater than $log_2 B$ bit.

3 PROBLEM 5

To find the normal equations solution, minimize

$$e(a,b) = \Sigma_i \left[(y_i - ax_i - b)^2\right]$$

Take the partial derivative of this equation:

$$\frac{\partial e(a,b)}{\partial a} = \Sigma_i \left[(y_i - ax_i - b)x_i\right]$$

$$\frac{\partial e(a,b)}{\partial b} = \Sigma_i \left[(y_i - ax_i - b)\right]$$

To find the minimal, let $\frac{\partial e(a,b)}{\partial a} = \frac{\partial e(a,b)}{\partial b} = 0$ Then,

$$\begin{cases} \Sigma_i(x_i^2)a + \Sigma_i x_i b & -\Sigma_i x_i y_i = 0 \\ \Sigma_i x_i a + nb & -\Sigma_i y_i = 0 \end{cases}$$

Solve this equation,

$$\begin{cases} a = \dfrac{n\Sigma_i(x_i y_i) - (\Sigma_i x_i)(\Sigma_i y_i)}{n\Sigma_i(x_i^2) - (\Sigma_i x_i)^2} \\ b = \dfrac{\Sigma_i(y_i)(\Sigma_i x_i)^2 - \Sigma_i(x_i y_i)\Sigma_i(x_i)}{n\Sigma_i(x_i^2) - (\Sigma_i x_i)^2} \end{cases}$$

4 PROBLEM 7

Suppose two convex hulls are both linearly separable and intersected.

Because they intersect, there must exits a point that are inside both convex hulls.

Let $x = [x_1\ x_2\ ...\ x_n]$ be a point in the intersecting area, $\alpha^{(1)} = [\alpha_1\ \alpha_2\ ...\ \alpha_n]$ for this point represented with the first convex hull's vectors, and $\alpha^{(2)} = [\alpha_1\ \alpha_2\ ...\ \alpha_n]$ for this point represented with the second convex hull's vectors.

$$\alpha^{(1)}x = \alpha^{(2)}x$$

Because these two convex hulls are linearly seperable, then we should have:
$\alpha^{(1)}x > 0$ if x is within the first convex hull
and
$\alpha^{(1)}x < 0$ if x is within the second convex hull
This conflicts with the first equation. So two convex hulls cannot be both linearly separable and intersected.

5 PROBLEM 8

$$\begin{aligned} & \nabla_A tr(ABA^TC) \\ =& \nabla_A tr(ABA^TC)(A^T\ treated\ as\ constant) + (\nabla_{A^T} tr(A^TCAB))^T(A\ treated\ as\ constant) \\ =& (BA^TC)^T + ((CAB)^T)^T \\ =& C^TAB^T + CAB \end{aligned}$$