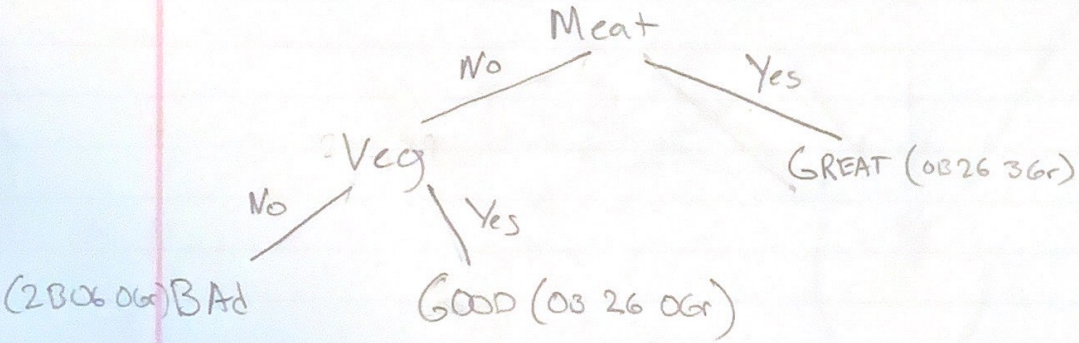
Decision Tree Learning HW

	Meat (N, Y)	Crust (D, S, T)	Veg (N, Y)	Quality (B, G, Gr)
1	Y	T	N	Gr
2	N	D	N	B
3	N	S	Y	G
4	Y	S	Y	Gr
5	Y	D	N	G
6	Y	D	Y	Gr
7	N	T	Y	G
8	Y	D	N	G
9	N	T	N	B



Entropy(S) = $\frac{2}{4}\log_2\frac{2}{4} - \frac{4}{4}\log_2\frac{4}{4} - \log_2\frac{3}{4} = 1.53$

Meat: 4N (2B, 2G, 0Gr) 5Y (0B, 2G, 3Gr)

$I(S) = \frac{4}{9}\left(-\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} - \frac{0}{4}\log_2\frac{0}{4}\right) + \frac{5}{9}\left(-\frac{0}{5}\log_2\frac{0}{5} - \frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5}\right) = .98$

Gain(S, Meat) = 1.53 - .98 = .55

Crust: 3T (1B, 1G, 1Gr) 4D (1B 2G 1Gr) 2S (0B 1G 1Gr)

$I(S) = \frac{3}{9}\left(-\frac{1}{3}\log_2\frac{1}{3} - \frac{1}{3}\log_2\frac{1}{3} - \frac{1}{3}\log_2\frac{1}{3}\right) + \frac{4}{9}\left(\frac{1}{4}\log_2\frac{1}{4} - \frac{2}{4}\log_2\frac{2}{4} - \frac{1}{4}\log_2\frac{1}{4}\right) + \frac{2}{9}\left(\frac{0}{2}\log_2\frac{0}{2} - \frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right) = 1.41$

Gain(S, crust) = 1.53 - 1.41 = 0.12

Veg 5N (2B 2G 1GR) 4Y (0B 2G 2GR)

$I(S) = \frac{5}{9}\left(-\frac{2}{5}\log_2\frac{2}{5} - \frac{2}{5}\log_2\frac{2}{5} - \frac{1}{5}\log_2\frac{1}{5}\right) + \frac{4}{9}\left(\frac{0}{4}\log_2\frac{0}{4} - \frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4}\right) =$

$.85 + .45 = 1.30$

Gain(S, Veg) = 1.53 - 1.30 = 0.23

Best Gain is Meat (.55 > .23 > .12)

No
At Meat level — 2Bud 2Good 0Gr

Entropy(S) = $\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log\frac{2}{4} - \frac{0}{4}\log_2\frac{0}{4} = 1$

Crust 2T (1B 1G 0Gr) 1D (1B 0G 0Gr) 1S (0B 1G 0Gr)

$I(S) = \frac{2}{4}\left(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right) + \frac{1}{4}\left(-\frac{1}{1}\log_2\frac{1}{1}\right) + \frac{1}{4}\left(-\frac{1}{1}\log_2\frac{1}{1}\right)$

$.5 + .125 + .125 = .75$

Gain(S, Crust) = 1 - .75 = .25

Veg: 2N (2B 0G 0Gr) 2Y (0B 2G 0Gr)

$I(S) = \frac{2}{4}\left(-\frac{2}{2}\log_2\left(\frac{2}{2}\right)\right) + \frac{2}{4}\left(-\frac{2}{2}\log_2\frac{2}{2}\right)$

$.25 + .25 = .5$

Gain(S, Crust) = 1 - .5 = .5

Vegetable is next best layer (.5 > .25)