Exercises for Logistic Regression and Decision Trees

Jordan Boyd-Graber
Digging into Data

February 25, 2013

1 Logistic Regression Formula

$$P(Y = 0|X) = \frac{1}{1 + \exp\left[w_0 + \sum_i w_i X_i\right]} \tag{1}$$

$$P(Y = 1|X) = \frac{\exp\left[w_0 + \sum_i w_i X_i\right]}{1 + \exp\left[w_0 + \sum_i w_i X_i\right]} \tag{2}$$

2 Example Weights

feature	symbol	weight
bias	w_0	0.1
"viagra"	w_1	2.0
"mother"	w_2	-1.0
"work"	w_3	-0.5
"nigeria"	w_4	3.0

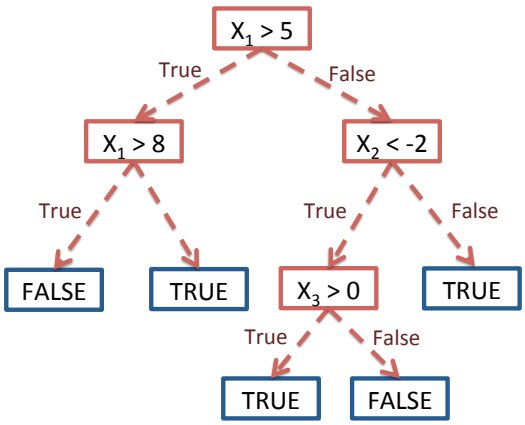
3 Example "Documents"

1. $X = \{\}$

2. $X = \{\text{Mother}, \text{Nigeria}\}$

3. $X = \{\text{Mother}, \text{Work}, \text{Viagra}, \text{Mother}\}$

4 Evaluating a Decision Tree



1. $(X_1, X_2, X_3) = (1, 1, 1)$
2. $(X_1, X_2, X_3) = (10, -3, 0)$

5 Creating Decision Trees

1. X AND Y (both must be true)

2. X OR Y (either can be true)

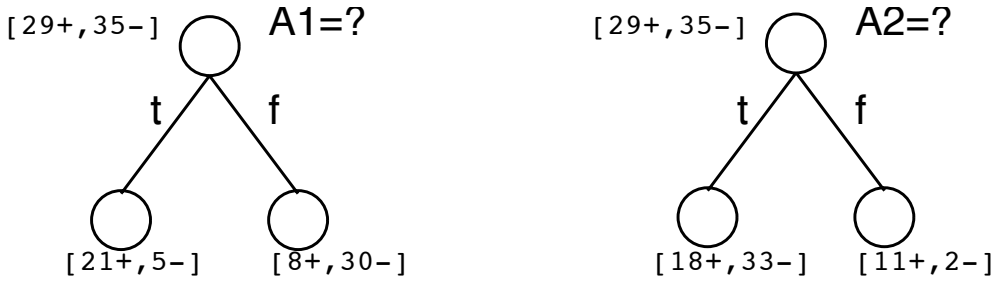
3. X XOR Y (one and only one is true)

6 Inducing Decision Tree Rules

$$\text{Entropy}(S) \equiv -p_{\oplus} \log_2 p_{\oplus} - p_{\ominus} \log_2 p_{\ominus} \tag{3}$$

$\text{Gain}(S, A)$ = expected reduction in entropy due to sorting on A

$$\text{Gain}(S, A) \equiv \text{Entropy}(S) \; - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{4}$$

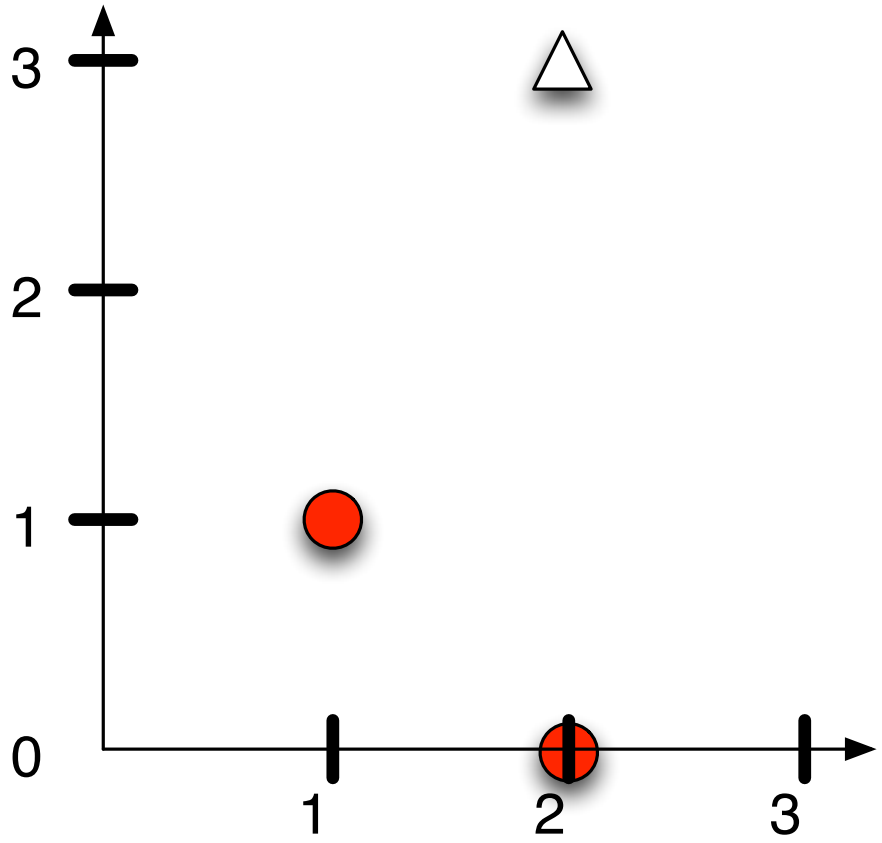


1. Entropy(S)

2. Gain(G, A_1)

3. Gain(G, A_2)

7 Support Vector Machines



1. Draw where the decision boundary should be
2. (Bonus) Find the equation for the line

8 Exponential Function Table

$\exp(-4.500)$	$= 0.011$	$\exp(-4.400)$	$= 0.012$	$\exp(-4.300)$	$= 0.014$	$\exp(-4.200)$	$= 0.015$
$\exp(-4.100)$	$= 0.017$	$\exp(-4.000)$	$= 0.018$	$\exp(-3.900)$	$= 0.020$	$\exp(-3.800)$	$= 0.022$
$\exp(-3.700)$	$= 0.025$	$\exp(-3.600)$	$= 0.027$	$\exp(-3.500)$	$= 0.030$	$\exp(-3.400)$	$= 0.033$
$\exp(-3.300)$	$= 0.037$	$\exp(-3.200)$	$= 0.041$	$\exp(-3.100)$	$= 0.045$	$\exp(-3.000)$	$= 0.050$
$\exp(-2.900)$	$= 0.055$	$\exp(-2.800)$	$= 0.061$	$\exp(-2.700)$	$= 0.067$	$\exp(-2.600)$	$= 0.074$
$\exp(-2.500)$	$= 0.082$	$\exp(-2.400)$	$= 0.091$	$\exp(-2.300)$	$= 0.100$	$\exp(-2.200)$	$= 0.111$
$\exp(-2.100)$	$= 0.122$	$\exp(-2.000)$	$= 0.135$	$\exp(-1.900)$	$= 0.150$	$\exp(-1.800)$	$= 0.165$
$\exp(-1.700)$	$= 0.183$	$\exp(-1.600)$	$= 0.202$	$\exp(-1.500)$	$= 0.223$	$\exp(-1.400)$	$= 0.247$
$\exp(-1.300)$	$= 0.273$	$\exp(-1.200)$	$= 0.301$	$\exp(-1.100)$	$= 0.333$	$\exp(-1.000)$	$= 0.368$
$\exp(-0.900)$	$= 0.407$	$\exp(-0.800)$	$= 0.449$	$\exp(-0.700)$	$= 0.497$	$\exp(-0.600)$	$= 0.549$
$\exp(-0.500)$	$= 0.607$	$\exp(-0.400)$	$= 0.670$	$\exp(-0.300)$	$= 0.741$	$\exp(-0.200)$	$= 0.819$
$\exp(-0.100)$	$= 0.905$	$\exp(0.000)$	$= 1.000$	$\exp(0.100)$	$= 1.105$	$\exp(0.200)$	$= 1.221$
$\exp(0.300)$	$= 1.350$	$\exp(0.400)$	$= 1.492$	$\exp(0.500)$	$= 1.649$	$\exp(0.600)$	$= 1.822$
$\exp(0.700)$	$= 2.014$	$\exp(0.800)$	$= 2.226$	$\exp(0.900)$	$= 2.460$	$\exp(1.000)$	$= 2.718$
$\exp(1.100)$	$= 3.004$	$\exp(1.200)$	$= 3.320$	$\exp(1.300)$	$= 3.669$	$\exp(1.400)$	$= 4.055$
$\exp(1.500)$	$= 4.482$	$\exp(1.600)$	$= 4.953$	$\exp(1.700)$	$= 5.474$	$\exp(1.800)$	$= 6.050$
$\exp(1.900)$	$= 6.686$	$\exp(2.000)$	$= 7.389$	$\exp(2.100)$	$= 8.166$	$\exp(2.200)$	$= 9.025$
$\exp(2.300)$	$= 9.974$	$\exp(2.400)$	$= 11.023$	$\exp(2.500)$	$= 12.182$	$\exp(2.600)$	$= 13.464$
$\exp(2.700)$	$= 14.880$	$\exp(2.800)$	$= 16.445$	$\exp(2.900)$	$= 18.174$	$\exp(3.000)$	$= 20.086$
$\exp(3.100)$	$= 22.198$	$\exp(3.200)$	$= 24.533$	$\exp(3.300)$	$= 27.113$	$\exp(3.400)$	$= 29.964$
$\exp(3.500)$	$= 33.115$	$\exp(3.600)$	$= 36.598$	$\exp(3.700)$	$= 40.447$	$\exp(3.800)$	$= 44.701$
$\exp(3.900)$	$= 49.402$	$\exp(4.000)$	$= 54.598$	$\exp(4.100)$	$= 60.340$	$\exp(4.200)$	$= 66.686$
$\exp(4.300)$	$= 73.700$	$\exp(4.400)$	$= 81.451$	$\exp(4.500)$	$= 90.017$	$\exp(4.600)$	$= 99.484$

9 Logarithm Table (Base 2)

$\lg(0.010) = -6.644$	$\lg(0.020) = -5.644$	$\lg(0.030) = -5.059$	$\lg(0.040) = -4.644$
$\lg(0.050) = -4.322$	$\lg(0.060) = -4.059$	$\lg(0.070) = -3.837$	$\lg(0.080) = -3.644$
$\lg(0.090) = -3.474$	$\lg(0.100) = -3.322$	$\lg(0.110) = -3.184$	$\lg(0.120) = -3.059$
$\lg(0.130) = -2.943$	$\lg(0.140) = -2.837$	$\lg(0.150) = -2.737$	$\lg(0.160) = -2.644$
$\lg(0.170) = -2.556$	$\lg(0.180) = -2.474$	$\lg(0.190) = -2.396$	$\lg(0.200) = -2.322$
$\lg(0.210) = -2.252$	$\lg(0.220) = -2.184$	$\lg(0.230) = -2.120$	$\lg(0.240) = -2.059$
$\lg(0.250) = -2.000$	$\lg(0.260) = -1.943$	$\lg(0.270) = -1.889$	$\lg(0.280) = -1.837$
$\lg(0.290) = -1.786$	$\lg(0.300) = -1.737$	$\lg(0.310) = -1.690$	$\lg(0.320) = -1.644$
$\lg(0.330) = -1.599$	$\lg(0.340) = -1.556$	$\lg(0.350) = -1.515$	$\lg(0.360) = -1.474$
$\lg(0.370) = -1.434$	$\lg(0.380) = -1.396$	$\lg(0.390) = -1.358$	$\lg(0.400) = -1.322$
$\lg(0.410) = -1.286$	$\lg(0.420) = -1.252$	$\lg(0.430) = -1.218$	$\lg(0.440) = -1.184$
$\lg(0.450) = -1.152$	$\lg(0.460) = -1.120$	$\lg(0.470) = -1.089$	$\lg(0.480) = -1.059$
$\lg(0.490) = -1.029$	$\lg(0.500) = -1.000$	$\lg(0.510) = -0.971$	$\lg(0.520) = -0.943$
$\lg(0.530) = -0.916$	$\lg(0.540) = -0.889$	$\lg(0.550) = -0.862$	$\lg(0.560) = -0.837$
$\lg(0.570) = -0.811$	$\lg(0.580) = -0.786$	$\lg(0.590) = -0.761$	$\lg(0.600) = -0.737$
$\lg(0.610) = -0.713$	$\lg(0.620) = -0.690$	$\lg(0.630) = -0.667$	$\lg(0.640) = -0.644$
$\lg(0.650) = -0.621$	$\lg(0.660) = -0.599$	$\lg(0.670) = -0.578$	$\lg(0.680) = -0.556$
$\lg(0.690) = -0.535$	$\lg(0.700) = -0.515$	$\lg(0.710) = -0.494$	$\lg(0.720) = -0.474$
$\lg(0.730) = -0.454$	$\lg(0.740) = -0.434$	$\lg(0.750) = -0.415$	$\lg(0.760) = -0.396$
$\lg(0.770) = -0.377$	$\lg(0.780) = -0.358$	$\lg(0.790) = -0.340$	$\lg(0.800) = -0.322$
$\lg(0.810) = -0.304$	$\lg(0.820) = -0.286$	$\lg(0.830) = -0.269$	$\lg(0.840) = -0.252$
$\lg(0.850) = -0.234$	$\lg(0.860) = -0.218$	$\lg(0.870) = -0.201$	$\lg(0.880) = -0.184$
$\lg(0.890) = -0.168$	$\lg(0.900) = -0.152$	$\lg(0.910) = -0.136$	$\lg(0.920) = -0.120$
$\lg(0.930) = -0.105$	$\lg(0.940) = -0.089$	$\lg(0.950) = -0.074$	$\lg(0.960) = -0.059$
$\lg(0.970) = -0.044$	$\lg(0.980) = -0.029$	$\lg(0.990) = -0.014$	$\lg(1.000) = 0.000$

10 Entropy Table (Base 2)

$$H(x) \equiv -x \lg(x) \tag{5}$$

$H(0.010)$	$= 0.066$	$H(0.020)$	$= 0.113$	$H(0.030)$	$= 0.152$	$H(0.040)$	$= 0.186$
$H(0.050)$	$= 0.216$	$H(0.060)$	$= 0.244$	$H(0.070)$	$= 0.269$	$H(0.080)$	$= 0.292$
$H(0.090)$	$= 0.313$	$H(0.100)$	$= 0.332$	$H(0.110)$	$= 0.350$	$H(0.120)$	$= 0.367$
$H(0.130)$	$= 0.383$	$H(0.140)$	$= 0.397$	$H(0.150)$	$= 0.411$	$H(0.160)$	$= 0.423$
$H(0.170)$	$= 0.435$	$H(0.180)$	$= 0.445$	$H(0.190)$	$= 0.455$	$H(0.200)$	$= 0.464$
$H(0.210)$	$= 0.473$	$H(0.220)$	$= 0.481$	$H(0.230)$	$= 0.488$	$H(0.240)$	$= 0.494$
$H(0.250)$	$= 0.500$	$H(0.260)$	$= 0.505$	$H(0.270)$	$= 0.510$	$H(0.280)$	$= 0.514$
$H(0.290)$	$= 0.518$	$H(0.300)$	$= 0.521$	$H(0.310)$	$= 0.524$	$H(0.320)$	$= 0.526$
$H(0.330)$	$= 0.528$	$H(0.340)$	$= 0.529$	$H(0.350)$	$= 0.530$	$H(0.360)$	$= 0.531$
$H(0.370)$	$= 0.531$	$H(0.380)$	$= 0.530$	$H(0.390)$	$= 0.530$	$H(0.400)$	$= 0.529$
$H(0.410)$	$= 0.527$	$H(0.420)$	$= 0.526$	$H(0.430)$	$= 0.524$	$H(0.440)$	$= 0.521$
$H(0.450)$	$= 0.518$	$H(0.460)$	$= 0.515$	$H(0.470)$	$= 0.512$	$H(0.480)$	$= 0.508$
$H(0.490)$	$= 0.504$	$H(0.500)$	$= 0.500$	$H(0.510)$	$= 0.495$	$H(0.520)$	$= 0.491$
$H(0.530)$	$= 0.485$	$H(0.540)$	$= 0.480$	$H(0.550)$	$= 0.474$	$H(0.560)$	$= 0.468$
$H(0.570)$	$= 0.462$	$H(0.580)$	$= 0.456$	$H(0.590)$	$= 0.449$	$H(0.600)$	$= 0.442$
$H(0.610)$	$= 0.435$	$H(0.620)$	$= 0.428$	$H(0.630)$	$= 0.420$	$H(0.640)$	$= 0.412$
$H(0.650)$	$= 0.404$	$H(0.660)$	$= 0.396$	$H(0.670)$	$= 0.387$	$H(0.680)$	$= 0.378$
$H(0.690)$	$= 0.369$	$H(0.700)$	$= 0.360$	$H(0.710)$	$= 0.351$	$H(0.720)$	$= 0.341$
$H(0.730)$	$= 0.331$	$H(0.740)$	$= 0.321$	$H(0.750)$	$= 0.311$	$H(0.760)$	$= 0.301$
$H(0.770)$	$= 0.290$	$H(0.780)$	$= 0.280$	$H(0.790)$	$= 0.269$	$H(0.800)$	$= 0.258$
$H(0.810)$	$= 0.246$	$H(0.820)$	$= 0.235$	$H(0.830)$	$= 0.223$	$H(0.840)$	$= 0.211$
$H(0.850)$	$= 0.199$	$H(0.860)$	$= 0.187$	$H(0.870)$	$= 0.175$	$H(0.880)$	$= 0.162$
$H(0.890)$	$= 0.150$	$H(0.900)$	$= 0.137$	$H(0.910)$	$= 0.124$	$H(0.920)$	$= 0.111$
$H(0.930)$	$= 0.097$	$H(0.940)$	$= 0.084$	$H(0.950)$	$= 0.070$	$H(0.960)$	$= 0.057$
$H(0.970)$	$= 0.043$	$H(0.980)$	$= 0.029$	$H(0.990)$	$= 0.014$	$H(1.000)$	$= 0.000$