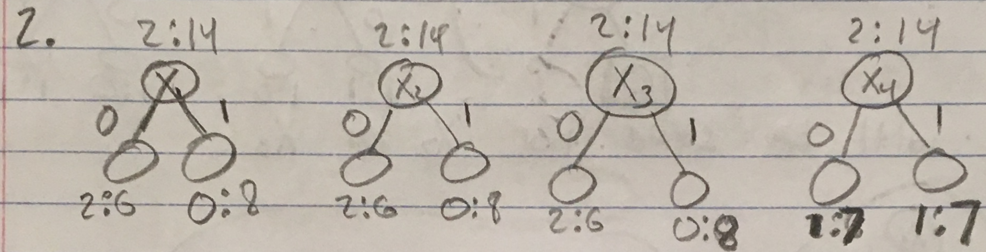
Jake Schinto

Homework 2

Decision Trees

1. with d attributes and 2 possible inputs per attribute, there are 2^d total unique inputs. Every input has two possible unique functions that return 0 or 1. So there are 2^{2^d} unique functions.

2.



Error: $2/16 = 1/8$ ↵ ↵ ↵

↳ $(1/4)/2 + 0/2 = 1/8$ ↵ ↵ $(1/8)/2 + (1/8)/2 = 1/8$

Entropy: $-1/8 \log_2(1/8) - 7/8 \log_2(7/8) = 0.544$

↳ $-1/4 \log_2(1/4) + 3/4 \log_2(3/4) = 0.811$ ← 0.544

↳ $0.811/2 = 0.406$ $((0.544)/2) \cdot 2 = 0.544$

Gain: 0.138 0

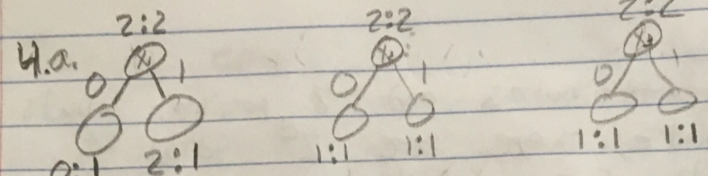
According to training error all nodes are equally good to be the root because none of them reduce error. By looking at entropy you actually can see that X_1, X_2, and X_3 all are good picks for the root, while there is still no gain from X_4.

3. a. The tree would have to be full, so every output would have to be explored. For this you would need 2^d input.

b. you would miss out on $2^{d!/k!}$ results, so error would be $2^{d!/k!}/2^d = 2^{(d!/k!)-d}$

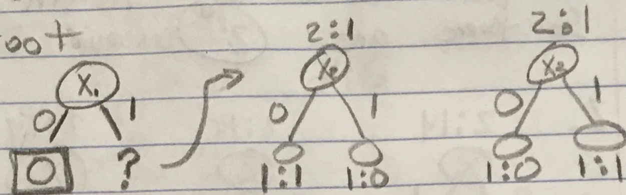
This is because each level you remove is 2^h, and that adds error that wasn't there before.

4.a.

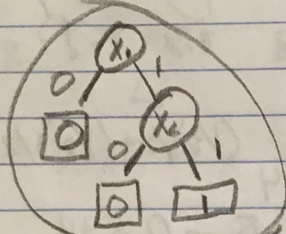


Gain: 0.082 0 0

X_1 is the root

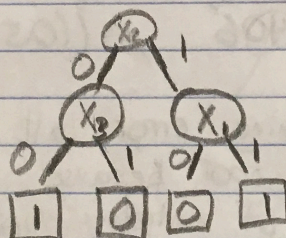


Gain will be same for X_2 or X_3



1/4 training error

b.



0 training error

More Probability

1. $E[X] = \sum_x x P(X=x) = (0)\cdot P(X=0) + (1)\cdot P(X=1)$
   $= P(X=1) = P(A)$

2. a. $P(X \cap Y) \stackrel{?}{=} P(X)\cdot P(Y)$

$P(X\cap Y) = 1/10 + 4/45 = 0.1889$ $\quad P(X) = 0.3889 \quad P(Y) = 0.4667$

$P(X)\cdot P(Y) = 0.1815 \neq 0.1889$ They are not independent

b. $P((X\cap Y)|Z) \stackrel{?}{=} P(X|Z)\cdot P(Y|Z)$

$$\frac{P(X\cap Y\cap Z)}{P(Z)} \stackrel{?}{=} \frac{P(X\cap Z)}{P(Z)}\cdot\frac{P(Y\cap Z)}{P(Z)}$$

$P(X \cap Y \cap Z) = 4/45 \qquad P(X \cap Z) = 2/9 \qquad P(Y \cap Z) = 4/15$

$P(Z) = 2/3$

$\frac{(4/45)}{(2/3)} = 0.1333 \qquad \frac{(2/9)}{(2/3)} \cdot \frac{(4/15)}{(2/3)} = 0.1333$

They are equal, so Yes X is conditionally dependant on Y Given Z.

c. $P(X=0 \mid X+Y>0) = \frac{P(X=0 \cap X+Y>0)}{P(X+Y>0)}$

$P(X+Y>0) = 1/10 + 1/10 + 1/15 + 2/15 + 8/45 + 4/45 = 2/3$

$P(X=0 \cap X+Y>0) = 1/10 + 8/45 = 5/18$

$\frac{(5/18)}{(2/3)} = 5/12 = 0.4167$

<u>MLE and MAP of Gaussians</u>

1. $\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n = \frac{1}{n}\sum_{i=1}^{n} x_i$

$MLE = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu_{ML})^2$

2.

$N(\mu; \mu_0, \sigma_0) \times N(x; \mu, \sigma) \qquad N = \frac{1}{\sqrt{2\pi}\sigma} \exp^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$

$\left(\frac{1}{\sqrt{2\pi}\sigma_0} \exp^{-\frac{1}{2}(\mu-\mu_0)^2/\sigma_0^2}\right) \times \left(\frac{1}{\sqrt{2\pi}\sigma} \exp^{-1/2(x-\mu)^2/\sigma^2}\right)$

$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}(\mu-\mu_0)^2/\sigma_0^2 - \frac{1}{2}(x-\mu)^2/\sigma^2\right\}$

$\exp \to -\frac{1}{2\sigma_0^2}(\mu^2 - 2\mu\mu_0 + \mu_0^2) - \frac{1}{2\sigma^2}(\mu^2 - 2x\mu + x^2)$

$\exp \to -\frac{1}{2\sigma^2}(2\mu^2 - 2\mu\mu_0 - 2x\mu + x^2 + \mu_0^2)$

$2\mu(\mu - \mu_0 - x) + x^2 + \mu_0^2 - 2x\mu_0 + 2x\mu_0$

$2\mu^2 - 2\mu\mu_0 - 2x\mu + (x - \mu_0)^2 + 2x\mu_0$

$2(\mu^2 - \mu\mu_0 - x\mu + x\mu_0) + (x - \mu_0)^2$

$2(\mu(\mu - \mu_0) - x(\mu - \mu_0)) + (x - \mu_0)^2$

$2(\mu - \mu_0)(\mu - x) + (x - \mu_0)^2$

$$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2}\left(2(\mu-\mu_0)(\mu-x)+(x-\mu_0)^2\right)/\sigma^2 \right\}$$

3. $MAP = N(\mu; \mu_N, \sigma_N^2)$

$$\mu_N = \frac{\sigma^2}{n\sigma_0^2+\sigma^2}\mu_0 + \frac{n\sigma_0^2}{n\sigma_0^2+\sigma^2}\mu_{ML} \leftarrow \mu_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}$$

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2}(\mu - \mu_N)/\sigma_N^2 \right\} = N(\mu; \mu_N, \sigma_N^2) = MAP$$

<u>MLE and MAP of Multinomial</u>

1. $\mu_{ML} = \frac{\sum a_{ik}}{n}$

$$MLE = \frac{1}{n-1}\sum_{i=1}^{n}(a_i - \mu_{ML})^2$$

2. $Dir(\theta | a+m) =$ posterior

$$\frac{\Gamma(a_0+n)}{\Gamma(a_1+m_1)\cdots\Gamma(a_k+m_k)} \cdot \prod_{k=1}^{K} \theta_k^{a_k+m_k-1}$$

where $\Gamma(x)$ is defined where $a_0 = \sum_{k=1}^{K} a_k$

$Dir(\mu | a+m) = MAP$

$$\frac{\Gamma(a_0+n)}{\Gamma(a_1+m_1)\cdots\Gamma(a_k+m_k)} \cdot \prod_{k=1}^{K} \mu_k^{a_k+m_k-1}$$

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u}\, du$$