HW #1

1) KL-divergence $KL(p\|q) = -\sum_x p(x)\log_2\left\{\frac{q(x)}{p(x)}\right\}$

$$I(x,y) = KL(p(x,y)\,\|\,p(x)p(y))$$

a) show that this definition of Mutual information is equivallent to the one given in class.

i.e. show $I(x,y) = H(x) - H(x|Y)$

$$H = -\sum_{i=1}^{n} P(x=i)\log_2(P(x=i)) \qquad H(x|Y) = \sum_Y P(Y=y)\,H(x|Y=y)$$

$$KL(p(x,y)\|p(x)p(y)) = -\sum_x p(x,y)\log_2\left\{\frac{p(x)p(y)}{p(x,y)}\right\}$$

$$= \sum_x \left\{p(x,y)\left[\log_2\left(\frac{p(x)}{p(x,y)}\right) + \log_2\left(\frac{p(y)}{p(x,y)}\right)\right]\right\}$$

for a joint distribution $p(x,y) = P(X=x|Y=y)\,P(Y=y) = P(Y=y|X=x)\cdot P(X=x)$

$$KL(p(x,y)\|p(x)p(y)) = -\sum_{x,y}\left\{p(x,y)\left[\log_2\left(\frac{P(X=x)}{P(Y=y|X=x)\cdot P(X=x)}\right) + \log_2\left(\frac{P(Y=y)}{P(X=x|P(Y=y))\,P(Y=y)}\right)\right.\right.$$

$$= -\sum_{x,y} p(x,y)\left\{\log_2(p(x)) + \log_2(p(y)) - \log_2(p(x,y))\right\}$$

marginalize over Y

$$= \underbrace{-\sum_x p(x)\log_2(p(x))}_{} + \sum_{x,y} p(x,y)\log\left\{\frac{p(y)}{p(x,y)}\right\}$$

$$= H(x) + \sum_{x,y} p(x,y)\log\left\{\frac{1}{P(X=x|Y=y)}\right\}$$

$$= H(x) + \sum_{x,y} p(x,y)\log\{P(X=x|Y=y)\} = H(x) - \underbrace{\sum_{Y,x,Y} P(X=x|Y=y)\,P(Y=y)\log_2\{P(x=x|Y=y)\}}_{\underbrace{\sum_Y P(Y=y)\sum_x P(X=x|Y=y)\log_2\{\ \}}_{H(x|Y)}}$$

$$I(x,y) = H(x) - H(x|Y)$$

∎

b) $I(x,y)$ is minimized when $H(X|Y) = H(X)$ i.e. X, Y are independent

(b2)

for a continuous x

$$H(x) = -\int p(x) \log(p(x))\, dx$$

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

derive entropy:

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

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

$$H(x) = -\frac{\log\left(\frac{1}{\sqrt{2\pi\sigma}}\right)}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} \exp\left\{ \frac{-(x-\mu)^2}{2\pi\sigma^2} \right\} dx + \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{2\pi\sigma^2} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

$$= \frac{1}{2}\left\{ \log\left(\frac{1}{\sqrt{2\pi\sigma}}\right) + 1 \right\}$$

(12.1) Bayes' Rule & point estimate

let D be the ~~pr~~ ~~de~~ D=+ indicates a person has the disease

let T denote the test T=+ denotes a positive test

$P(T=+ \mid D=+) = 0.95$ $\quad P(D=+) = 0.05$

$P(T=+ \mid D=-) = 0.05$

$$P(T=+) = \sum_D P(T=+ \mid D) \cdot P(D) = (0.95)(0.05) + (0.05)(1-0.05) = 0.095$$

we want $P(D=+ \mid T=+) = \frac{P(T=+ \mid D=+)\,P(D=+)}{P(T=+)} = \frac{(0.95)(0.05)}{(0.095)} = \frac{(0.95)(0.05)}{2(0.95)(0.05)} = \frac{1}{2}$

(12.2) *

(3.1) we must write the entropy for X where X can be + or −

we have $N(+)$ ~~&~~ $N(-)$

$$H(x) = -\sum_{x=i}^{n} P(x=i)\log_2(P(x=i))$$

here ~~xx~~ $n=2$ $x = \oplus$ or $\ominus$

$$H(x) = -\left\{P(x=+)\log_2(P(x=+)) + P(x=-)\log_2(P(x=-))\right\}$$

to do a top down approach suppose we are at a node we treat it

like it has no children and prune it ~~xx~~ if improves.

In this way we implement the recursive call after the pruning

for a bottom-up approach we travel down to a leaf so we implement the recursive call before pruning.

top down pruning results in 348 nodes w/ a test correct of 87%

bottom up prunning results in 512 nodes w/ a test correction at 87%

with a smaller ε we keep more nodes but seem to have a small to no increase in correctness