Refined by PMF/CDF
PDF

Mass

$X \sim f(x) \longrightarrow P(X = x)$

Density

$X \sim f(x) \quad F(x) \rightarrow P(X \geq x)$

- in domain
- $\forall x \in X,\ P(x) \geq 0$
- $\int P(x)\,dx = 1$

Joint

$P(X, Y) \rightarrow P(X = x, Y = y)$

Marginal

$P(x) \rightarrow P(X = x, Y = \text{arbitrary}) = \sum_y P(X, Y = y)$

or

$\int P(x, y)\,dy$

Conditional

$P(X \mid Y = y) \rightarrow \dfrac{P(X, Y = y)}{P(Y = y)}$

Do Bayes Rule here

Chain Rule immediately follows!

$$P(a, b, c, \ldots) = \underbrace{\frac{P(a, b, c)}{P(b, c)}}_{P(a \mid b c)} \cdot \frac{P(b, c)}{P(c)} \cdot P(c)$$

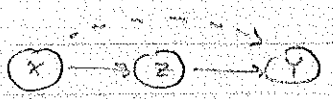
$P(a \mid bc) \cdot P(b \mid c) \cdot P(c)$

Independence:

Share no information $\quad P(a, b) = P(a \mid b) \cdot P(b)$ $\quad a \perp b$

($P(a \mid b) = P(a)$)

CI - Shows no information given you know z.



$X \perp Y \mid Z$

$$P(X,Y|Z) = \frac{P(X,Y,Z)}{P(Y,Z)} \cdot \frac{P(Y,Z)}{P(Z)}$$

$$P(X|Y,Z) \cdot P(Y|Z)$$

no info gained → $P(X|Z)$

Examples:

0.05	0.1	0.2
0.3	0.15	0.2

0.35
0.65

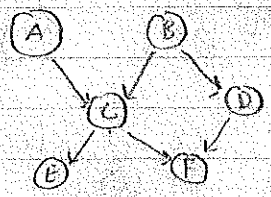
0.15	0.28	0.57
0.46	0.23	0.31

Introduce class

0.35	0.25	0.4

5	10	20
30		20

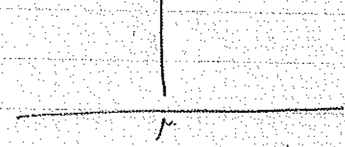
0.15	0.4	0.5
0.85	0.6	0.5

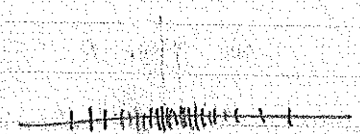


$$P(F|C,D) \cdot P(E|C) \cdot P(D|B) \cdot P(C|A,B) \cdot P(B) \cdot P(A)$$

See Bishop 8.5

Dirac Delta (Empirical Dist.





$Y=0 \to P$, $Y=1 \to (1-P)$

Mixture: $Y \sim \text{Bernoulli}(P)$

$(X \mid Y=y) \sim N(y\mu_1 + (1-y)\mu_2, 1)$

CDF of X: $P(X \le x) = \sum_{y=0}^{1} P(X \le x \mid Y=y)\,P(Y=y)$

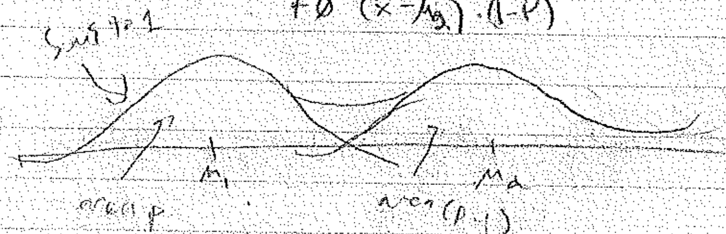
$F_X(x) =$

$= P(X \le x \mid Y=0)P(Y=0) + P(X \le x \mid Y=1)P(Y=1)$

$= P(X-\mu_1 \le x-\mu_1 \mid Y=0)P(Y=0) + P(X-\mu_2 \le x-\mu_2 \mid Y=1)P(Y=1)$

$= \Phi(x-\mu_1)P(Y=0) + \Phi(x-\mu_2)P(Y=1)$

$\frac{d}{dx} F_X(x) \quad \phi(x-\mu_1)\cdot P + \phi(x-\mu_2)\cdot(1-P)$



Y can be latent / Prior

CLT: Let $X_1, X_2, \ldots, X_n$ iid with mean μ and var $\sigma^2 < \infty$

$$\sqrt{n}(\bar{X} - \mu) \Rightarrow N(0, \sigma^2)$$

LLN:

$$\bar{X}_n \xrightarrow{p} \mu.$$

$$\mathrm{Var}(f(x)) = \mathbb{E}\left((f(x) - \mathbb{E}(f(x)))^2\right)$$

$$\mathrm{Cov}(f(x), g(x)) = \mathbb{E}\left((f(x) - \mathbb{E}(f(x)))(g(y) - \mathbb{E}(g(y)))\right)$$

$$\mathbb{E}(x) = \frac{1}{n}\sum_n x \quad \text{or} \quad \int x \cdot p_x \, dx$$

$$\frac{1}{n}\sum_n (x - \mu)^2$$

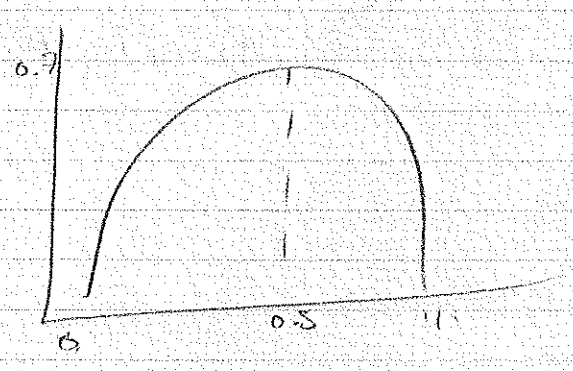
$$\frac{1}{n}\sum_n (x - \mu_x)(y - \mu_y)$$

Useful functions

Sigmoid $\sigma(x) = \frac{1}{1+e^{-x}}$ Probs, Bern, Saturates

Softplus $\zeta(x) = \log(1 + \exp(x))$

Softened version of ReLU



$$H(x) = E_{x \sim p}[I(x)] = -E_{x \sim p}[\log P(x)]$$

Sunny	Cloudy	rainy	foggy
00	01	10	11

need 2 bits → motivates $\log_2\left(\frac{1}{p}\right)$

If Sunny more likely save bits

1/2	1/4	1/8	1/8
0	10	110	111

$\log(2) = 1$ $\log(4) = 2$ $\log(8) = 3$

$\frac{1}{2}$ $\frac{1}{4}$ $\frac{1}{8} \cdot \frac{1}{8}$

$\frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{2}{8} \times 3 = 1\frac{3}{8}$

other $\frac{1}{2} \cdot 2 = \frac{1}{2} \cdot 4 = 2$

$$D_{KL}(P\|Q) = \mathbb{E}_{x\sim P}\left[\log \frac{P(x)}{Q(x)}\right] = \mathbb{E}_{x\sim P}[\log P(x) - \log(Q)]$$

$$-\sum P \ln\left(\frac{P}{q}\right) \geq -\sum P\left(\frac{q}{p} - 1\right)$$

by $\ln x \leq x - 1$

$$-\sum P\left(\frac{q}{p} - 1\right) = -\sum q + \sum p = 0$$

$$H(P,Q) = -\mathbb{E}_{x\sim P}\log(Q(x))$$

how much it cost to code Q under P's coding scheme.

$$H(P,Q) = H(P) + D_{KL}(P\|Q)$$

minimizing is equiv.