- Discuss Least squares: $z=Xw+n$
    - $J=||z-Xw||^2 = z'z-2z'Xw+w'X'Xw$; $X=U\Sigma V'$
    - $w=(X'X)^{-1}X'z = V\Sigma^{\dagger}U'z$; $\dagger \Rightarrow$ pseudo inverse
    - zhat=UU'z =Pz…. projection
    - z-zhat=(I-UU')z=(I-P)z
    - $X=QR$ …. $w=R^{-1}Q'z$….. Q via Gram-Schmidt, Householder, Givens
- Discuss R(X), N(X'), R(X'), N(X)
- Discuss what happens when singular value is small.

- Ways to overcome: Ridge, L1, Elastic net, PCR, PLS, probabilistic PCA,…..

$$J(\underline{w}) = \| \underline{z} - X\underline{w} \|_2^2 + \mu \| \underline{w} \|_2^2 \text{ .....}\textit{assume input and output data is centered}$$

$$\Rightarrow \underline{\hat{w}} = \left(X^T X + \mu I_p \right)^{-1} X^T \underline{z} = X^T \left(X X^T + \mu I_N \right)^{-1} \underline{z} \text{ ....via Matrix Inversion Lemma}$$

$$(or)\, \underline{\hat{w}} = X^T \alpha; \underline{\alpha} = \left(X X^T + \mu I_N \right)^{-1} \underline{z} \text{....}\textit{Kernel form} \Rightarrow \hat{z} = \underline{\hat{w}}^T \underline{x} = \sum_{i=1}^{N} \alpha_i \underline{x}_i^T \underline{x}$$

$$so,\ \underline{\hat{z}} = \underbrace{X \left(X^T X + \mu I_p \right)^{-1} X^T}_{P_\mu} \underline{z} = \sum_{j=1}^{p} \left(\frac{\lambda_j^2}{\lambda_j^2 + \mu} \right) \underline{u}_j \underline{u}_j^T \underline{z}; \lambda_j^2 = \textit{eigen value of } X^T X = \sigma_j^2$$

$$J(\underline{w}) = \| \underline{z} - X\underline{w} \|_2^2 + \mu \| \underline{w} \|_1$$

$$J(\underline{w}) = \| \underline{z} - X\underline{w} \|_2^2 + \mu_1 \| \underline{w} \|_1 + \mu_2 \| \underline{w} \|_2^2$$

$$X_r = X V_r = U_r \Sigma_r \text{....}\textit{feature selection}; \textit{data transformation / reduction}$$

$$V_r = \text{first } r \text{ singular vectors corresponding to largest } r \text{ singular values}$$

$$\hat{w} = (X_r^T X_r)^{-1} X_r^T \underline{z} = \sum_{i=1}^{r} \left(\frac{\underline{u}_i^T \underline{z}}{\sigma_i} \right) \underline{v}_i$$

PLS and Probabilistic PCA in lecture 10….. Feature selection

- Bayes' Rule … draw graph
- a statistical procedure that describes the optimal way to update one's beliefs when making new observations (i.e., receiving new sensory input or evidence).

$$P(A|B)=\frac{P(A,B)}{P(B)}=\frac{P(B|A)P(A)}{P(B)}=\frac{P(B|A)P(A)}{\sum_{A'}P(B|A')P(A')}$$

- Sum Rule (Total Probability Theorem)

$$P(B)=\sum_{A}P(A,B)$$

- Product Rule

$$P(A,B)=P(B|A)P(A)=P(A|B)P(B)$$

valid for continuous and hybrid contexts.

$$\text{Re}\,gression:\ p(\underline{w}|\underline{z},X)=\frac{p(\underline{z}|\underline{w},X)\,p(\underline{w})}{p(\underline{z}|X)}=\frac{p(\underline{z}|\underline{w},X)\,p(\underline{w})}{\int_{\underline{w}}p(\underline{z}|\underline{w},X)\,p(\underline{w})d\,\underline{w}}$$

$$Classification:\ P(A|\underline{z})=\frac{p(\underline{z}|A)P(A)}{\sum_{A'}p(z|A')P(A')}$$

$$\text{Re}\,gression\ with\ quantized\ observations:$$

$$p(\underline{w}|A)=\frac{P(A|\underline{w})\,p(\underline{w})}{\int_{\underline{w}}p(A|\underline{w})\,p(\underline{w})d\,\underline{w}}$$

B	C	P(A=1|B,C)	P(A=0|B,C)
0	0	0.10	0.90
0	1	0.99	0.01
1	0	0.80	0.20
1	1	0.25	0.75

P(B=1)=0.65

P(C=1)=0.77

$$P(B=1|A=0)=\frac{P(A=0|B=1)P(B=1)}{P(A=0)}=\frac{P(A=0|B=1)P(B=1)}{P(A=0|B=1)P(B=1)+P(A=0|B=0)P(B=0)}$$

$$P(A=0|B=1)=\sum_{C=0}^{1}P(A=0,C|B=1)=\sum_{C=0}^{1}P(A=0|B=1,C)P(C|B=1)$$

$$=\sum_{C=0}^{1}P(A=0|B=1,C)P(C)=0.20*0.23+0.75*0.77=0.6235$$

Similarly, $P(A=0 \mid B=0)=\sum_{C=0}^{1} P(A=0 \mid B=0,C)P(C)=0.90*0.23+0.01*0.77=0.2147$

$$P(B=1 \mid A=0)=\frac{0.6235*0.65}{0.6235*0.65+0.2147*0.35}=0.8436$$

$$P(B=0 \mid A=0)=1-P(B=1 \mid A=0)=0.1564$$

$$P(C=1 \mid A=0)=\frac{P(A=0 \mid C=1)P(C=1)}{P(A=0 \mid C=1)P(C=1)+P(A=0 \mid C=0)P(C=0)}$$

$$P(A=0 \mid C=1)=\sum_{B=0}^{1} P(A=0 \mid C=1,B)P(B)=0.01*0.35+0.75*0.65=0.4910$$

$$P(A=0 \mid C=0)=\sum_{B=0}^{1} P(A=0 \mid C=0,B)P(B)=0.90*0.35+0.20*0.65=0.4450$$

$$P(C=1 \mid A=0)=\frac{0.4910*0.77}{0.4910*0.77+0.4450*0.23}=0.7870$$

$$P(C=0 \mid A=0)=1-P(C=1 \mid A=0)=0.2130$$

$$P(C=0 \mid A=1)=\frac{P(A=1 \mid C=0)P(C=0)}{P(A=1 \mid C=0)P(C=0)+P(A=1 \mid C=1)P(C=1)}$$

$$=\frac{0.555*0.23}{0.555*0.23+0.509*0.77}=0.2457$$

- Discuss graphical models