

1) sigmoid $\sigma(x) = \frac{1}{1+e^{-x}}$, softplus $\zeta(x) = \ln(1+e^x)$

a) (i) $(1-\sigma(x)) = \frac{1-e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}} = \frac{e^{-x}}{1+e^{-x}} \Rightarrow \frac{e^{-x}}{1+e^{-x}} \cdot \frac{e^{-x}}{e^{-x}} = \frac{1}{1+e^{-x}}$

$$\Rightarrow (1-\sigma(x)) = \frac{1}{1+e^{-x}} = \frac{1}{1+e^{-(x)}} = \sigma(-x)$$

(ii) $\frac{d}{dx} \sigma(x) = \frac{d}{dx} \left[\frac{1}{1+e^{-x}} \right] = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{e^{-x} - e^{-x}}{(1+e^{-x})^2} = \frac{e^{-x}}{(1+e^{-x})^2}$

$$= \frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{1+e^{-x}} \Rightarrow \frac{d\sigma(x)}{dx} = \frac{1}{1+e^{-x}} \cdot \left[1 - \frac{1}{1+e^{-x}} \right] = \sigma(x)(1-\sigma(x)) = \sigma(x)\sigma(-x)$$

(iii) $\ln(\sigma(x)) = \ln\left(\frac{1}{1+e^{-x}}\right) = \ln((1+e^{-x})^{-1}) = -\ln(1+e^{-x}) = -\zeta(-x)$

(iv) $\frac{d\zeta(x)}{dx} = \frac{d}{dx} [\ln(1+e^x)] = \frac{1}{1+e^x} \cdot e^x = \frac{e^x}{1+e^x} \cdot \frac{e^{-x}}{e^{-x}} = \frac{1}{1+e^{-x}} = \sigma(x)$

(v) $x = \ln(e^x) = \ln\left(e^x \cdot \frac{1+e^{-x}}{2+e^{-x}}\right) = \ln\left(\frac{1+e^x}{2+e^{-x}}\right) = \ln\left(\frac{\frac{1}{2+e^{-x}}}{\frac{1}{2+e^{-x}}}\right) = \ln\left(\frac{\sigma(x)}{\sigma(-x)}\right) = \ln\left(\frac{\sigma(x)}{2-\sigma(x)}\right)$

(vi) $x = \ln(e^x) = \ln(e^x + 1 - 1) = \ln\left(e^{\ln(e^x + 1)} - 1\right) = \ln\left(e^{\zeta(x)} - 1\right)$

(vii') $\zeta(x) - \zeta(-x) = \ln(1+e^{-x}) - \ln(1+e^x) \stackrel{\text{use (iii)}}{\Rightarrow} -\ln(\sigma(-x)) + \ln(\sigma(x)) = \zeta(x) - \zeta(-x)$

$$\Rightarrow -\ln(\sigma(-x)) + \ln(\sigma(x)) = \ln\left(\frac{\sigma(x)}{\sigma(-x)}\right) \stackrel{\text{use (iv)}}{\Rightarrow} x = \zeta(+x) - \zeta(-x)$$

2) (i) $f(\underline{w}) = \sum_{n=1}^N (z_n - \underline{w}^T \underline{x}_n)^2 \rightarrow \nabla_{\underline{w}} f(\underline{w}) = -2 \sum_{n=1}^N (z_n - \underline{w}^T \underline{x}_n) \underline{x}_n$

 $\nabla_{\underline{w}}^2 f(\underline{w}) = \nabla_{\underline{w}} [\nabla_{\underline{w}} f(\underline{w})] \rightarrow ???.$
 $f(\underline{w}) = \sum_{n=1}^N (z_n^2 - 2z_n \underline{w}^T \underline{x}_n + (\underline{w}^T \underline{x}_n)^2) = \sum_{n=1}^N (z_n^2 - 2z_n \underline{w}^T \underline{x}_n + \underline{w}^T \underline{x}_n \underline{x}_n^T \underline{w})$
 $\nabla_{\underline{w}}^2 f(\underline{w}) = 2 \sum_{n=1}^N (2z_n \underline{x}_n + 2\underline{w}^T \underline{x}_n \underline{x}_n^T \underline{w}) \quad * \text{Don't understand this definition}$
 $\nabla_{\underline{w}}^2 f(\underline{w}) = \sum_{n=1}^N \left(\frac{\partial}{\partial \underline{w}} [\underline{w}^T \underline{x}_n \underline{x}_n^T \underline{w}] \right) = 2 \sum_{n=1}^N \underline{x}_n \underline{x}_n^T > 0 \rightarrow \text{convex!}$

$x_n x_n^T \sim \|x_n\|^2 > 0$
for vectors (real)

(ii) $f(\underline{w}) = \sum_{n=1}^N (z_n - y_n)^2, y_n \equiv \sigma(\underline{w}^T \underline{x}_n) = \frac{1}{e^{-\underline{w}^T \underline{x}_n} + 1}$

 $\nabla_{\underline{w}} f(\underline{w}) = \sum_{n=1}^N -2(z_n - y_n) y_n \underline{x}_n = -2 \sum_{n=1}^N (z_n - y_n) y_n (1-y_n) \underline{x}_n = \nabla_{\underline{w}} f(\underline{w})$
 $\nabla_{\underline{w}}^2 f(\underline{w}) = -2 \sum_{n=1}^N [-y_n(1-y_n) \cdot y_n(1-y_n) + (z_n - y_n)(y_n(1-y_n)(1-y_n) - y_n y_n(1-y_n))] \underline{x}_n \underline{x}_n^T$
 $\nabla_{\underline{w}}^2 f(\underline{w}) = 2 \sum_{n=1}^N [y_n(1-y_n)(y_n(1-y_n) - (z_n - y_n)(1-2y_n))] \underline{x}_n \underline{x}_n^T, \quad y_n \geq 0, z_n \geq 0$

$\nabla_{\underline{w}}^2 f(\underline{w})$ is not convex since $\exists z_n, y_n$ such that $(+) (+) (+) (+) - (\frac{z_n - y_n}{y_n}) (\frac{z_n - y_n}{y_n}) = \frac{z_n - y_n}{y_n} (\frac{z_n - y_n}{y_n}) < 0$, $\frac{z_n - y_n}{y_n} < 0 \Rightarrow \nabla_{\underline{w}}^2 f(\underline{w}) < 0$

(iii) $f(\underline{w}) = -\sum_{n=1}^N [z_n \log y_n + (1-z_n) \log(1-y_n)] \rightarrow \text{assume } \log = \ln$

 $\nabla_{\underline{w}} f(\underline{w}) = -\sum_{n=1}^N [z_n \cdot \frac{y_n(1-y_n)}{y_n} + (1-z_n) \cdot -\frac{y_n(1-y_n)}{1-y_n}] \underline{x}_n$
 $\nabla_{\underline{w}} f(\underline{w}) = -\sum_{n=1}^N [z_n - y_n] \underline{x}_n$
 $\nabla_{\underline{w}}^2 f(\underline{w}) = \sum_{n=1}^N y_n(1-y_n) \underline{x}_n \underline{x}_n^T > 0 \rightarrow \text{convex}$

$$P(x)P(y|x) = \\ P(y)P(x|y)$$

3) (labeled 2)

a) $P(A, B, C) = P(A)P(B|A)P(C|B)$, $P(A=0) = 0.3$, $P(A=1) = 0.7$

$P(C=1|B) = 1 - P(C=0|B)$, $P(B=2|A) = 1 - P(B=0|A) - P(B=1|A)$

$\sum_{A,B,C} P(A, B, C) = 7 \quad \checkmark$

$P(0, 0, 0) = 0.3 \cdot 0.3 \cdot 0.2 = 0.018$, $P(0, 0, 1) = 0.3 \cdot 0.3 \cdot 0.8 = 0.072$

$P(0, 1, 0) = 0.3 \cdot 0.2 \cdot 0.2 = 0.012$, $P(0, 1, 1) = 0.3 \cdot 0.2 \cdot 0.8 = 0.048$

$P(1, 0, 0) = 0.3 \cdot 0.5 \cdot 0.4 = 0.060$, $P(1, 0, 1) = 0.3 \cdot 0.5 \cdot 0.6 = 0.090$

$P(1, 1, 0) = 0.7 \cdot 0.1 \cdot 0.2 = 0.014$, $P(1, 1, 1) = 0.7 \cdot 0.1 \cdot 0.8 = 0.056$

$P(1, 2, 0) = 0.7 \cdot 0.4 \cdot 0.2 = 0.056$, $P(1, 2, 1) = 0.7 \cdot 0.4 \cdot 0.8 = 0.224$

$P(2, 0, 0) = 0.7 \cdot 0.5 \cdot 0.4 = 0.140$, $P(2, 0, 1) = 0.7 \cdot 0.5 \cdot 0.6 = 0.210$

$P(B) = \sum_{A,C} P(A, B, C) \Rightarrow P(B=0) = 0.16$, $P(B=1) = 0.34$, $P(B=2) = 0.5$

$P(C) = \sum_{A,B} P(A, B, C) \Rightarrow P(C=0) = 0.3$, $P(C=1) = 0.7$

$P(A, C) = \sum_B P(A)P(B|A)P(C|B) = \sum_B P(A) \frac{P(B)P(A|B)}{P(A)} P(C|B) \quad \text{WR}$

$P(A=0, C=0) = 0.090$, $P(A=0, C=1) = 0.210$, $P(A=1, C=0) = 0.210$, $P(A=1, C=1) = 0.490$

From cursory glance, one sees that $P(A, C) = P(A)P(C)$. However, this is in general false. We know B depends on A & C depends on B therefore C and A are not mutually disjoint.

$P(A|C=1) = ?$

$P(B|C) = \frac{P(B)P(C|B)}{P(C)}$

$\Rightarrow P(B|C) = \frac{P(B)P(C=1|B)}{P(C=1)}$

$\rightarrow P(B=1|C=1) = P(B=1) \cdot P(C=1|B=1) = \frac{0.34 \cdot 0.8}{0.8} \approx 0.389 = P(B=1|C=1)$

$P(A|C) = \sum_B P(A|B)P(B|C) = \sum_B \frac{P(A)P(B|A)}{P(B)} \cdot \frac{P(B)P(C|B)}{P(C)} = \sum_B \frac{P(A)}{P(C)} \cdot P(B|A)P(C|B)$

$P(A=0|C=1) = 0.3$, $P(A=1|C=1) = 0.7$

b) Binomial distribution \rightarrow

$$(\text{i}) \text{ One draw: } P(B) = \frac{B}{K}, P(W) = \frac{K-B}{K}$$

$$\rightarrow P(B=n_B) = \binom{N}{n_B} \left(\frac{B}{K}\right)^{n_B} \left(\frac{K-B}{K}\right)^{N-n_B}$$

(ii) ~~$E(n_B)$~~ $\rightarrow E(n_B) =$

$$E(n_B) = N \cdot \frac{B}{K}$$

$$\text{Var}(n_B) = N \cdot \frac{B}{K} \left(1 - \frac{B}{K}\right)$$

$$\rightarrow N=5, B=2, K=10 \rightarrow P(B)=0.2, P(W)=0.8, E[n_B]=1, \text{Var}[n_B]=0.8$$

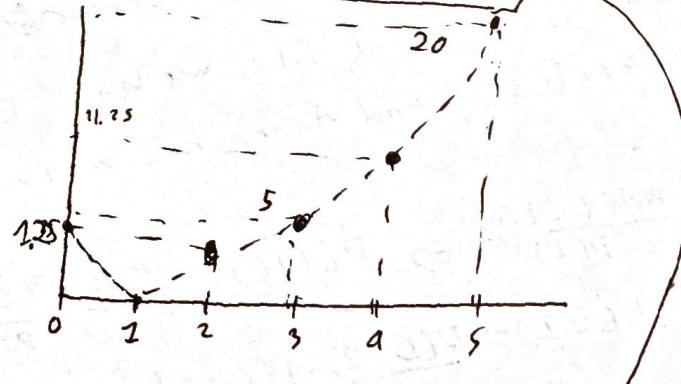
$$\rightarrow N=400, B=2, K=10 \rightarrow P(B)=0.2, P(W)=0.8, E[n_B]=80, \text{Var}[n_B]=64$$

$$\begin{aligned} (\text{iii}) f_B &= B/K, x = \frac{(n_B - f_B N)^2}{N f_B (1-f_B)} \rightarrow E(x) = \sum_{n_B=0}^N n_B \frac{(n_B - f_B N)^2}{N f_B (1-f_B)} / \sum_{n_B=0}^N \frac{(n_B - f_B N)^2}{N f_B (1-f_B)} \\ E(x) &= \sum_{n_B=0}^N \frac{1}{N f_B (1-f_B)} \sum_{n_B=0}^N [n_B^3 - 2n_B^2 f_B N + n_B f_B^2 N^2] / \sum_{n_B=0}^N \frac{n_B^2}{N f_B (1-f_B)} \\ &= \left[\frac{1}{N f_B (1-f_B)} \left[\frac{n^2 (n+1)^2}{4} - f_B \frac{n^2 (n+1)(2n+1)}{3} + f_B^2 \frac{n^3 (n+1)}{2} \right] \right] = E(x) \end{aligned}$$

$$= \frac{1}{N f_B (1-f_B)} \left[\frac{6n^2 + 12n^3 + 6n^4}{4} - f_B \frac{n^2 (n+1)(2n+1)}{3} + f_B^2 \frac{n^3 (n+1)}{2} \right] = E(x)$$

$$N=5, f_B=2/5$$

$$E(x) = \frac{\sum_{i=0}^5 n_i \frac{(n_i - 1)^2}{(2-1/5)}}{\sum_{i=0}^5 \frac{(n_i - 1)^2}{4/5}} \approx 4.2$$



$$\hookrightarrow E(x) = \frac{\frac{n^2 (n+1)^2}{4} - f_B \frac{n^2 (n+1)(2n+1)}{3} + f_B^2 \frac{n^3 (n+1)}{2}}{\frac{N(N+1)(2N+1)}{6} - f_B \frac{n^2 (n+1)}{2} + f_B^2 N^3}$$

(c) (labelled b)
~~Exponential~~ $\rightarrow P(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{else} \end{cases}$ up to ∞
 but bounded by 1 cm to 15cm

~~Exponential~~
 $\rightarrow P(x) = \begin{cases} ce^{-\lambda x} & 1 \leq x \leq 15 \\ 0 & \text{else} \end{cases}$

$$\rightarrow 1 = \int_{-\infty}^{\infty} P(x) dx = \int_{-15}^{15} ce^{-\lambda x} dx = \frac{-c}{\lambda} [e^{-\lambda \cdot 15} - e^{-\lambda}] = 1 \Rightarrow c = \frac{\lambda}{e^{-\lambda} - e^{-15\lambda}}$$

| | |
|---|--|
| $P(x) = \begin{cases} \frac{1}{e^{-\lambda} - e^{-15\lambda}} e^{-\lambda x} & 1 \leq x \leq 15 \\ 0 & \text{else} \end{cases}$ | $E[x] = \int_{-\infty}^{\infty} x P(x) dx$ |
| $\text{Var}(x) = \int_{-\infty}^{\infty} (x - E[x])^2 P(x) dx = \frac{1}{\lambda^2 [e^{-\lambda} - e^{-15\lambda}]}$ | $E[x] = \frac{1}{\lambda [e^{-\lambda} - e^{-15\lambda}]}$ |

4) (labelled 2) $B(\theta; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} \quad \theta \in [0, 1]$
 $a > 0, b > 0$ $\sim \text{Beta}(a, b)$

$$E[\theta] = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^\infty \theta \theta^{a-1} (1-\theta)^{b-1} d\theta = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \theta^a (1-\theta)^{b-1} d\theta = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)}{\Gamma(a+b+1)}$$
 $= \frac{\Gamma(a)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+1)} \xrightarrow{\text{Ratio Test}} \frac{a}{a+b} = E[x]$

$$\text{Var}[x] = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 (\theta - \frac{a}{a+b})^2 \theta^{a-1} (1-\theta)^{b-1} d\theta = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \left(\theta^2 - \frac{2a}{a+b} \theta + \frac{a^2}{(a+b)^2} \right) \theta^{a-1} (1-\theta)^{b-1} d\theta$$
 $= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} - \frac{2a}{a+b} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} + \frac{a^2}{(a+b)^2} \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \right)$
 $= \frac{a(a+1)}{(a+b+1)(a+b)} - \frac{2a}{a+b} \frac{a}{a+b} + \frac{a^2}{(a+b)^2} = \frac{a^3 + a^2 b + a^2 + b^2 a - a^3 - a^2 b - a^2}{a(a+1)(a+b)} = \frac{ab}{(a+b+1)(a+b)^2} = \text{Var}(x)$

$$\text{Mode}(x) \rightarrow \frac{\partial B}{\partial \theta} = 0 = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} ((a-1)\theta^{a-2} (1-\theta)^{b-2} + (b-1)\theta^{a-2} (1-\theta)^{b-2})$$

assume $a, b > 1$;
 ~~$(a-1)\theta^{a-2} (1-\theta)^{b-2}$~~ $\frac{(a-1)\theta^{a-2} (1-\theta)^{b-2}}{(b-1)\theta^{a-2} (1-\theta)^{b-2}} = 1 \rightarrow \frac{a-1}{b-1} \theta^{a-2} (1-\theta)^{b-2} = 1 \rightarrow (a-1)(1-\theta) = (b-1)\theta \rightarrow \frac{a-1}{a+b-2}$

$$\text{Mode}(\theta) = \frac{a-1}{a+b-2}$$

$$\text{Entropy } H(\theta) = -\int p(x) \ln(p(x)) dx$$

$$H(\theta) = -\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \theta^{a-1} (1-\theta)^{b-1} \left(\ln\left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right) + \ln(\theta^{a-1}) + \ln((1-\theta)^{b-1}) \right) d\theta$$

$$H(\theta) = -\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cancel{\ln(\Gamma(a)\Gamma(b))} \int_0^1 \theta^{a-1} (1-\theta)^{b-1} ((a-1)\ln(\theta) + (b-1)\ln(1-\theta)) d\theta$$

$$H(\theta) = -\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \ln\left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right) - \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \theta^{a-1} (1-\theta)^{b-1} ((a-1)\ln(\theta) + (b-1)\ln(1-\theta)) d\theta$$

$$H(\theta) = -\ln\left(\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\right) - \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 (1-\theta)^{b-1} ((a-1)\ln(\theta) + (b-1)\ln(1-\theta)) d\theta$$

5) (labelled q) $p(x) = N(\mu, \sigma^2), q(x) = N(\nu, \delta^2)$

$$KL(P||q) = \int_{-\infty}^{\infty} p(x) \ln\left(\frac{p(x)}{q(x)}\right) dx = \frac{1}{2} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left[\frac{-(x-\mu)^2}{2\sigma^2} + \frac{(\mu-\nu)^2}{2\delta^2} + \ln\left(\frac{\delta}{\sigma}\right) \right] dx = -\frac{1}{2} \ln(2\pi) - \ln(\sigma) + \frac{1}{2} \ln(2\pi) + \ln(\delta)$$

$$\frac{-\sqrt{2\pi}\sigma}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2\sigma^2}} u^2 du = \frac{-\sqrt{2\pi}\sigma}{\sqrt{2\pi}\sigma} \cdot \frac{2!}{1!} \cdot \left(\frac{\pi}{2}\right)^3 \cdot 2 = \frac{1}{2}$$

$$u = \frac{x-\mu}{\sqrt{2\sigma}}, du = \frac{dx}{\sqrt{2\sigma}}, x = \sqrt{2\sigma}u + \mu$$

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ln\left(\frac{\delta}{\sigma}\right) dx = \underline{\ln\left(\frac{\delta}{\sigma}\right)}$$

$$\frac{\sqrt{2\sigma}}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2\sigma^2}} \frac{(\sqrt{2\sigma}u + \mu - \nu)^2}{2\delta^2} du = \frac{1}{2\delta^2\sqrt{2}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2\sigma^2}} \underbrace{\frac{2\sigma^2 u^2}{2\sigma^2 u^2 + 2\sqrt{2\sigma}u(\mu-\nu) - \mu\nu + \mu^2 + \nu^2}}_{\text{odd} \Rightarrow 0} du$$

$$= \frac{\sqrt{\pi} \cdot 2}{2\delta^2\sqrt{\pi}} \left[2\sigma^2 \cdot 2 \cdot \left(\frac{1}{2}\right)^3 + \frac{1}{2}(\mu^2 + \nu^2 - \mu\nu) \right] = \underline{\frac{\sigma^2 + \mu^2 + \nu^2 - \mu\nu}{2\delta^2}}$$

$$KL(P||q) = \ln \frac{\delta}{\sigma} - \frac{1}{2} + \frac{\sigma^2 + \mu^2 + \nu^2 - \mu\nu}{2\delta^2} = \frac{1}{2} \left[\ln \frac{\delta^2}{\sigma^2} - 1 + \frac{(\mu-\nu)^2}{\delta^2} + \frac{\sigma^2}{\delta^2} \right]$$

Multivariate:
 $\downarrow \text{Don't fully understand how to convert vector form from scale}$

$$KL(P||q) = \frac{1}{2} \left[\ln \frac{\det(\Sigma)}{\det(\sigma)} + (\mu - \nu)^T \Sigma^{-1} (\mu - \nu) + \text{tr}(\Sigma^{-1} \sigma) - n \right]$$

$$P(x) = \sum_{i=1}^k \alpha_i N(\underline{\mu}_i, \Sigma_i), \quad \sum_{i=1}^k \alpha_i = 1, \quad \alpha_i \geq 0$$

$$hL(P||q) = \sum_{i=1}^k \alpha_i \left[\frac{1}{2} \left(\ln \frac{\det(S)}{\det(\Sigma_i)} + (\underline{\mu}_i - \underline{\nu})^T S^{-1} (\underline{\mu}_i - \underline{\nu}) + \text{tr}(S^{-1} \Sigma_i) - n \right) \right]$$

If $q(x)$ is a weighted gaussian it becomes difficult to compute as is related to the $-\ln(q(x))$ which can quickly become complicated once there are many gaussians correlated with each other.

$$6) (4S) I(x; Y|z) = H(x|z) - H(x|Y|z), \text{ where } N(\mu, \sigma^2), Y \sim N(\nu, \delta^2), z \sim N(m, s^2)$$

$$P\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = N\left(\begin{pmatrix} \mu_x \\ \mu_y \\ \mu_z \end{pmatrix}; \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_z^2 \end{pmatrix}\right) = N\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}; \begin{pmatrix} \sigma_x^2 & f_{xy}\sigma_x\sigma_y & f_{xz}\sigma_x\sigma_z \\ f_{yx}\sigma_x\sigma_y & \sigma_y^2 & f_{yz}\sigma_y\sigma_z \\ f_{zx}\sigma_x\sigma_z & f_{zy}\sigma_y\sigma_z & \sigma_z^2 \end{pmatrix}\right)$$

$$P(x,y,z) = N\left(\begin{pmatrix} \mu_x \\ \mu_y \\ \mu_z \end{pmatrix}; \begin{pmatrix} \sigma_x^2 & f_{xy}\sigma_x\sigma_y & f_{xz}\sigma_x\sigma_z \\ f_{yx}\sigma_x\sigma_y & \sigma_y^2 & f_{yz}\sigma_y\sigma_z \\ f_{zx}\sigma_x\sigma_z & f_{zy}\sigma_y\sigma_z & \sigma_z^2 \end{pmatrix}\right), I(x; Y|z) = I(x; Y) - I(x; Y|z)$$

$$I(x; Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}} e^{-\frac{(z-\mu_z)^2}{2\sigma_z^2}} dz dy dx$$

$$I(x; Y) = H(x) + H(y) - H(x, y)$$

$$H(x) = \int_{-\infty}^{\infty} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} dx \rightarrow \text{d.d. E.P.S. i. f. r. a. f.}$$

$$\text{D.E.X.} \quad I(x; Y|z) = H(x|z) - H(x|Y, z) = H(x|z) - H(x|Y) - H(x|Y|z) ?$$

$$H(x|z) = - \int_{-\infty}^{\infty} \frac{(x-\mu_x - \frac{\sigma_{xz}}{\sigma_z^2}(z-\mu_z))^2}{2(\sigma_x^2 - \frac{\sigma_{xz}^2}{\sigma_z^2})} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} dz \rightarrow \text{d.d. t. r. i. s. f. r. k. l.} \quad - \text{L. t. c. t. r. s. ??}$$

$$H(x|z) = \frac{1}{2} \left[1 + \ln(2\pi) + \ln\left(\sigma_x^2 - \frac{\sigma_{xz}^2}{\sigma_z^2}\right) \right] \quad | \quad \text{otaz}$$

$$H(x|Y, z) = - \int_{-\infty}^{\infty} \left(\sqrt{2\pi} \left(\sigma_x^2 - \sigma_{xy}^2 - \sigma_{xz}^2 \right) \begin{bmatrix} \sigma_y^2 & \sigma_{yz} \\ \sigma_{yx} & \sigma_z^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{xy} \\ \sigma_{xz} \end{bmatrix} \right)^{-1} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} dz \quad | \quad \text{a. (....) d(x)}$$

$$\rightarrow \text{K.L. c. n. s.} \rightarrow H(x|Y, z) = \frac{[\sigma_{xy} \sigma_{xz}]}{\sigma_y^2 \sigma_z^2 - \sigma_{yz}^2} \begin{bmatrix} \sigma_y^2 & \sigma_{yz} \\ \sigma_{yx} & \sigma_z^2 \end{bmatrix} \begin{bmatrix} \sigma_{xy} \\ \sigma_{xz} \end{bmatrix} = \frac{[\sigma_{xy} \sigma_{xz}]}{\sigma_y^2 \sigma_z^2 - \sigma_{yz}^2} \begin{bmatrix} \sigma_y^2 \sigma_{xy} + \sigma_{yz} \sigma_{xz} \\ \sigma_y^2 \sigma_{xz} - \sigma_{yz} \sigma_{xy} \end{bmatrix}$$

$$\frac{1}{2} \left(1 + \ln(2\pi) + \ln\left(\frac{\sigma_y^2 \sigma_{xz}^2 - \sigma_{xy} \sigma_{yz} \sigma_{xz} + \sigma_y^2 \sigma_{xz}^2 - \sigma_{xz} \sigma_{yz} \sigma_{xy}}{\sigma_y^2 \sigma_z^2 - \sigma_{yz}^2}\right) \right)$$

$$I(x; Y|z) = H(x|z) - H(x|Y, z)$$

$$= \frac{1}{2} \ln \left(\frac{\sigma_x^2 \sigma_{xy}^2 + \sigma_y^2 \sigma_{xz}^2 + 2\sigma_{xy} \sigma_{yz} \sigma_{xz}}{\sigma_x^2 \sigma_y^2 \sigma_z^2 - \sigma_x^2 \sigma_{yz}^2 - \sigma_z^2 \sigma_{xy}^2 - \sigma_y^2 \sigma_{xz}^2 + 2\sigma_{xy} \sigma_{yz} \sigma_{xz}} \right)$$

$$I(x; Y|z) = \frac{1}{2} \ln \left(\frac{\left(\sigma_x^2 - \frac{\sigma_{xz}^2}{\sigma_z^2} \right) \left(\sigma_y^2 \sigma_z^2 - \sigma_{yz}^2 \right)}{\sigma_x^2 \sigma_y^2 \sigma_z^2 - \sigma_x^2 \sigma_{yz}^2 - \sigma_z^2 \sigma_{xy}^2 - \sigma_y^2 \sigma_{xz}^2 + 2\sigma_{xy} \sigma_{yz} \sigma_{xz}} \right) \quad \underline{xz, yz, xy}$$

7) (Linear Regression) $J(\omega) = \frac{1}{N} \sum_{n=1}^N (z_n - \omega^T x_n)^2 = \frac{1}{N} \| z - X\omega \|_2^2$
 $\nabla_{\omega} J = \frac{2}{N} \sum_{n=1}^N (z_n - \omega^T x_n) x_n \Leftrightarrow \nabla_{\omega} J = \frac{2X^T}{N} (z - X\omega) = 0 = X^T z = X^T X \omega$
 $X = \begin{bmatrix} 1 & \dots & 1 \\ x_1 & \dots & x_N \end{bmatrix}$ $N \gg p$

b) $\hat{z} = X\hat{\omega} = X(X^T X)^{-1} X^T z = ((X^T X)^{-1})^T X^T z = \hat{z}$
 $\boxed{z = X(X^T X)^{-1} X^T z + X(X^T X)^{-1} X^T \nu}$
 $= X(X^T X)^{-1} (X^T z) + X(X^T X)^{-1} X^T \nu$
 $= X\hat{z} + X(X^T X)^{-1} X^T \nu = X\hat{z} + P\nu = \hat{z}$

c) $\hat{z} - \bar{z} = z - X\hat{\omega} = z - X(X^T X)^{-1} X^T z = I\nu = X\omega - X\hat{\omega} + \nu - P\nu = I\nu = I\nu - P\nu = \bar{z} - \hat{z}$

d) $E[J(\omega) | \omega = \hat{\omega}] = \frac{1}{N} \| z - X\hat{\omega} \|_2^2 = \frac{1}{N} E[\| z - \hat{z} \|_2^2]$
 $= \frac{1}{N} E[\| (I-P)\nu \|_2^2] = \frac{1}{N} E[((I-P)\nu)^T (I-P)\nu] = \frac{1}{N} E[\nu^T (I-P)^2 \nu]$
 $= \frac{1}{N} E[\nu^T \nu] \text{tr}[(I-P)^2] = \frac{\sigma^2}{N} (N-p) = \boxed{\sigma^2(1-p/N) = E[J(\omega)]}$
 $\xrightarrow{\text{MP}} \nu \sim N(0, \sigma^2 I_p)$ $\nu \sim N(0, \sigma^2 I_p)$ $\xrightarrow{\text{tr}(I-P)^2 = (I-P)(I-P)^T = \text{tr}(I-P) = N-p}$ $\xrightarrow{\frac{N-p}{N}}$

e) $z_{N+1} - \hat{z}_{N+1} = z_{N+1} - X^T \hat{\omega} = z_{N+1} - X^T X(X^T X)^{-1} X^T z$
 $= X_{N+1}^T \omega + \nu_{N+1} - X_{N+1}^T \hat{\omega} = X_{N+1}^T \omega + \nu_{N+1} - X_{N+1}^T (X^T X)^{-1} X^T (X\omega + \nu_{\text{real}})$
 $= \boxed{X_{N+1}^T \omega + \nu_{N+1} - X_{N+1}^T (X^T X)^{-1} (X^T X)\omega - X_{N+1}^T (X^T X)^{-1} \nu_{\text{real}}} = \boxed{V_{N+1} - X_{N+1}^T (X^T X)^{-1} V_{\text{real}}}$
 $= \boxed{X_{N+1}^T (\omega - X\omega) + \nu_{N+1}} = \boxed{V_{N+1} - X_{N+1}^T (X^T X)^{-1} V_{\text{real}} = z_{N+1} - \hat{z}_{N+1}}$

f) $E[(z_{N+1} - \hat{z}_{N+1})^2] = E[(V_{N+1} - X_{N+1}^T (X^T X)^{-1} V_{\text{real}})^2] = E[V_{N+1}^2 - 2V_{N+1} X_{N+1}^T (X^T X)^{-1} V_{\text{real}} + X_{N+1}^T (X^T X)^{-1} V_{\text{real}}^T (X^T X)^{-1} X_{N+1}]$
 $= E[V_{N+1}^2] - 2E[V_{N+1} X_{N+1}^T (X^T X)^{-1} X^T V_{\text{real}}] + E[X_{N+1}^T (X^T X)^{-1} X^T V_{\text{real}} V_{\text{real}}^T (X^T X)^{-1} X_{N+1}]$
 $= \sigma^2 - 2\sigma^2 \underbrace{\text{tr}[(X^T X)^{-1} X^T X]}_{= N-p} - 2 \cdot 0 + \text{tr}[X_{N+1}^T (X^T X)^{-1} X^T X (X^T X)^{-1} X_{N+1}] \xrightarrow{\sigma^2} \Sigma_{N+1}?$
 $= \sigma^2 + \sigma^2 \underbrace{\text{tr}[(X^T X)^{-1}]}_{= N-1} \xrightarrow{\sigma^2} \Sigma = \sigma^2 + \sigma^2 \underbrace{\text{tr}[(X^T X)^{-1}]}_{= N-1} = \sigma^2 (1 + \frac{p}{N-1}) \approx \sigma^2 (1 + \frac{p}{N}) \approx E[z_{N+1} - \hat{z}_{N+1}]$

$$g) \hat{z}_{-i} = \hat{\omega}^T \underline{x}_i = \left(\underbrace{((\underline{x}^T \underline{x})^{-1} \underline{x}^T)_{\underline{z}_i}}_{\text{L } i^{th} \text{ element of } P} \right)^T \underline{x}_i = \underline{z}_i^T \underbrace{\underline{x} (\underline{x}^T \underline{x})^{-1} \underline{x}_i}_{= w_i} = w_i$$

$$= \underline{z}_i^T P_{ii} ? = P_{ii} z_i \quad \text{AA No clue}$$

$$\tilde{z}_i = \underline{x} \hat{\omega} = \underline{x} w + P_{ii}$$

$$= \hat{\omega}^T \underline{x}_i + v$$

$$h) (z_i - \hat{z}_{-i})^2 = \left(z_i - \frac{\hat{z}_i - P_{ii} z_i}{1 - P_{ii}} \right)^2 = \left(\frac{z_i - P_{ii} z_i - \hat{z}_i + P_{ii} z_i}{1 - P_{ii}} \right)^2 = \left[\frac{(z_i - \hat{z}_i)^2}{1 - P_{ii}} \right] = (z_i - \hat{z}_i)^2$$

$$i) E[J_{Local}] = \frac{1}{N} \sum_{i=1}^N \frac{E[(z_i - \hat{z}_i)^2]}{(1 - P_{ii})^2}$$

$$= \boxed{\frac{\sigma^2}{N} \sum_{i=1}^N \frac{1 + P_{ii}}{(1 - P_{ii})^2} = E[J_{WOCV}]}$$

use $P_{ii} \neq 0$ $\Rightarrow E[J_{Local}] = \sigma^2 (1 - P_{ii})$
 expect $\text{tr}[I] = 1$
 & use the i^{th} component of P
 instead of $\text{tr}[P] \Rightarrow \text{tr}[P] = P_{ii}$ because it's discrete

The "static" method does well at low N which makes sense since you're only dealing w/ a few data points. So you can attack them one at a time. The "shuffled" method does well dealing with more at a time or average of data points rather than one at a time. This helps to avoid solving the regression for only one or two data points.