Paul Simmerling
ECE 6141

PS 1

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

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

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

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

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

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

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

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

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

(vii) $\zeta(x) - \zeta(-x) = \cancel{\ln(1+e^{x}) - \ln(1+e^{-x})} \xRightarrow{\text{use (iii)}} \boxed{-\ln(\sigma(-x)) + \ln(\sigma(x)) = \zeta(x) - \zeta(-x)}$

$$\Rightarrow -\ln(\sigma(-x)) + \ln(\sigma(x)) = \ln\left(\frac{\sigma(x)}{\sigma(-x)}\right) \xRightarrow{\text{use (v)}} \boxed{x = \zeta(+x) - \zeta(-x)}$$

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

$\nabla^2_{\underline{w}} f(\underline{w}) = \nabla_{\underline{w}} [\nabla_{\underline{w}} f(\underline{w})] \rightarrow ???$

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

$\cancel{\nabla f(\underline{w}) = \sum_{n=1}^{N} (-2 z_n \underline{x}_n + 2 \underline{w}^T \underline{x}_n \underline{x}_n) =}$

$\boxed{\nabla^2_{\underline{w}} f(\underline{w}) = \sum_{n=1}^{N} \left(\frac{\partial}{\nabla_{\underline{w}}} [\underline{w}_n^T \underline{x}_n \underline{x}_n^T \underline{w}_n] \right) = 2 \sum_{n=1}^{\infty} \underline{x}_n \underline{x}_n^T} \boxed{> 0 \rightarrow \text{convex!}}$

* Don't understand this definition

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

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

$\nabla_{\underline{w}} f(\underline{w}) = \sum_{n=1}^{N} 2 (z_n - y_n)(-y_n') \nabla_{\underline{w}} (\underline{w}^T \underline{x}_n) = \boxed{-2 \sum_{n=1}^{N} (z_n - y_n) y_n (1 - y_n) \underline{x}_n = \nabla_{\underline{w}} f(\underline{w})}$

$\nabla^2_{\underline{w}} f(\underline{w}) = -2 \sum_{n=1}^{N} [-y_n(1-y_n) \cdot y_n(1-y_n) + (z_n - y_n)(y_n(1-y_n)(1-y_n) - y_n y_n (1-y_n))]$

$\boxed{\nabla^2_{\underline{w}} f(\underline{w}) = 2 \sum_{n=1}^{N} [y_n(1-y_n)(y_n(1-y_n) - (z_n - y_n)(1 - 2y_n))] \underline{x}_n \underline{x}_n^T}$

$(+)(+)((+)(+) - (\overset{y_n > z_n}{\underset{+}{-}})(\underset{+}{-})), \quad z_n = 1, y_n = 0 \rightarrow \nabla^2_{\underline{w}} f < 0, \quad z_n = 1, y_n = 1 \ \nabla^2_{\underline{w}} f > 0$

$y_n \geq 0$
$z_n \geq 0$

$\boxed{\nabla^2_{\underline{w}} f(\underline{w}) \text{ is not convex for all } z_n, y_n}$

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

$\nabla_{\underline{w}} f(\underline{w}) = -\sum_{n=1}^{N} \left[z_n \cdot \frac{y_n(1-y_n)}{y_n} + (1 - z_n) \cdot \frac{-y_n(1-y_n)}{1 - y_n} \right] \underline{x}_n$

$\boxed{\nabla_{\underline{w}} f(\underline{w}) = -\sum_{n=1}^{N} [z_n - y_n] \underline{x}_n}$

$\boxed{\nabla^2_{\underline{w}} f(\underline{w}) = \sum_{n=1}^{N} y_n (1 - y_n) \underline{x}_n \underline{x}_n^T > 0 \rightarrow \text{convex}}$

3) (labeled 2)

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

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

$\underline{P(C=1|B) = 1 - P(C=0|B)}$, $\underline{P(B=2|A) = 1 - P(B=0|A) - P(B=1|A)}$

- $P(0,0,0) = 0.3\cdot0.3\cdot0.2 = 0.018$, $P(0,0,1) = 0.3\cdot0.3\cdot0.8 = 0.072$
- $P(0,1,0) = 0.3\cdot0.2\cdot0.2 = 0.012$, $P(0,1,1) = 0.3\cdot0.2\cdot0.8 = 0.048$
- $P(0,2,0) = 0.3\cdot0.5\cdot0.4 = 0.060$, $P(0,2,1) = 0.3\cdot0.5\cdot0.6 = 0.090$
- $P(1,0,0) = 0.7\cdot0.1\cdot0.2 = 0.014$, $P(1,0,1) = 0.7\cdot0.1\cdot0.8 = 0.056$
- $P(1,1,0) = 0.7\cdot0.4\cdot0.2 = 0.056$, $P(1,1,1) = 0.7\cdot0.4\cdot0.8 = 0.224$
- $P(1,2,0) = 0.7\cdot0.5\cdot0.4 = 0.140$, $P(1,2,1) = 0.7\cdot0.5\cdot0.6 = 0.210$

$\sum_{A,B,C} P(A,B,C) = 1$ ✓

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

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

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

$\underline{P(A=0,C=0) = 0.090,\ P(A=0,C=1) = 0.210,\ P(A=1,C=0) = 0.210,\ P(A=1,C=1) = 0.490}$

From a 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.

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

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

$\rightarrow P(B=1|C=1) = \frac{P(B=1)\cdot P(C=1|B=1)}{P(C=1)} = \frac{0.34\cdot0.8}{0.7} \approx 0.389 = P(B=1|C=1)$

$\underline{P(B=0|C=1) \approx 0.183}$, $\underline{P(B=2|C=1) \approx 0.429}$

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

$\underline{P(A=0|C=1) = 0.3}$, $\underline{P(A=1|C=1) = 0.7}$

b) Binomial distribution →

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

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

(ii) $\cancel{E[n_B] = n_B \frac{B}{K}}$ $\boxed{E(n_B) = N \cdot \frac{B}{K}}$

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

→ $N = 5$, $B = 2$, $K = 10$ → $P(B) = 0.2$, $P(W) = 0.8$, $E[n_B] = 1$, $Var[n_B] = 0.8$

→ $N = 400$, $B = 2$, $K = 10$ → $P(B) = 0.2$, $P(W) = 0.8$, $E[n_B] = 80$, $Var(n_B) = 64$

(iii) $f_B = B/K$, $x = \frac{(n_B - f_B N)^2}{N f_B (1 - f_B)}$ → $E(x) = \sum_{n_B=0}^{N} n_B \frac{(n_B - f_B N)^2}{N f_B (1 - f_B)} \Big/ \sum_{n_B=0}^{N} \frac{(n_B - f_B N)^2}{N f_B (1 - f_B)}$

$$E(x) = \cancel{\sum_{n_B=0}^{N} n_B} \frac{1}{N f_B (1 - f_B)} \sum_{n_B=0}^{N} \left[n_B^3 - 2 n_B^2 f_B N + n_B f_B^2 N^2\right] \Big/ \frac{1}{N f_B (1 - f_B)} \sum_{n_B=0}^{N} n_B^2 - 2 n_B f_B N + f_B^2 N^2$$

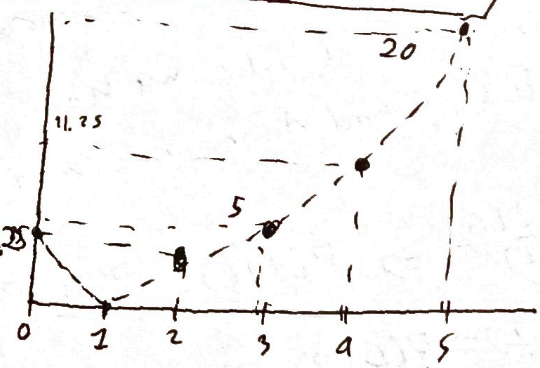
$\frac{N(N+1)}{2}$

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

$$= \cancel{\frac{1}{N f_B (1 - f_B)}} \left[\cancel{6N^4 + 12N^3 + 6N^2}\right.$$

$N = 5$, $f_B = 1/5$

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



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

(c) (labeled b)

exponential $\rightarrow p(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{else} \end{cases}$ ~~with λ=~~ but bounded by 1 cm to 15 cm

~~EDIT~~

$$\rightarrow p(x) = \begin{cases} c e^{-\lambda x} & 1 \leq x \leq 15 \\ 0 & \text{else} \end{cases}$$

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

$$\boxed{p(x) = \begin{cases} \frac{\lambda}{e^{-\lambda} - e^{-15\lambda}} e^{-\lambda x} & 1 \leq x \leq 15 \\ 0 & \text{else} \end{cases}}$$

$$\boxed{E[x] = \int_{-\infty}^{\infty} x\, p(x)\, dx \qquad E[x] = \frac{1}{\lambda[e^{-\lambda} - e^{-15\lambda}]}}$$

$$\boxed{Var(x) = \int_{-\infty}^{\infty} (x - E[x])^2 p(x)\, dx = \frac{1}{\lambda^2[e^{-\lambda} - e^{-15\lambda}]}}$$

---

4) (labeled 3) $B(\theta; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1} \qquad \theta \in [0,1]$

$a > 0,\ b > 0,\ \theta \sim B(a,b)$

$$E[\theta] = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_{-\infty}^{\infty} \theta\, \theta^{a-1}(1-\theta)^{b-1} d\theta = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \theta^{\overbrace{a}^{(a+1)-1}}(1-\theta)^{b-1} d\theta = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)}$$

$$= \frac{\Gamma(a+1)\Gamma(a+b)}{\Gamma(a)\Gamma(a+b+1)} \xrightarrow{\Gamma \text{ dif}} \boxed{\frac{a}{a+b} = E[x]}$$

$\Gamma(z+1) = z\Gamma(z)$

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

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{\Gamma(a+2)\Gamma(b)}{\Gamma(a+b+2)} - \frac{2a}{a+b} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} + \frac{a^2}{(a+b)^2} \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \right)$$

$$= \frac{a(a+1)}{(a+b+1)(a+b)} - \frac{2a}{a+b}\frac{a}{a+b} + \frac{a^2}{(a+b)^2} = \frac{\overbrace{a(a+1)(a+b)}^{a^3+a^2b+a^2+ba} \overbrace{- 2a^2(a+b+1) + a^2(a+b+1)}^{-a^3-a^2b-a^2}}{(a+b+1)(a+b)^2} = \boxed{\frac{ab}{(a+b+1)(a+b)^2} = Var(x)}$$

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

~~assume a,b>~~ assume $a, b > 1$:

$$\frac{(a-1)\theta^{a-2}(1-\theta)^{b-1}}{(b-1)\theta^{a-1}(1-\theta)^{b-2}} = 1 \rightarrow \frac{a-1}{b-1}\theta^{-1}(1-\theta) = 1 \rightarrow (a-1)(1-\theta) = (b-1)\theta \rightarrow \frac{a-1}{a+b-2}$$

$$\boxed{mode(\theta) = \frac{a-1}{a+b-2}}$$

Entropy $H(\theta) = -\int p(x)\ln(p(x))\,dx$

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

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

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

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

5) (labeled 4) $p(x) = N(\mu, \sigma^2)$, $q(x) = N(\nu, \delta^2)$

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

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

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

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

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

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

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

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

Multivariate:

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

* Don't fully understand how to convert to vector form from scale

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

$$\boxed{KL(p \| q) = \sum_{i=1}^{L} \alpha_i \left[\frac{1}{2} \left(\ln \frac{\det(S)}{\det(\sigma_i)} + (\underline{\mu}_i - \underline{\nu})^T S^{-1} (\underline{\mu}_i - \underline{\nu}) + tr(S^{-1} \sigma_i) - n \right) \right]}$$

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

6) (15) $I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$, $X \sim N(\mu, \sigma^2)$, $Y \sim N(\nu, \delta^2)$, $Z \sim N(m, s^2)$

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

$\cancel{P(x, y, z) = N\left(\begin{bmatrix} \mu_x \\ \mu_y \\ \mu_z \end{bmatrix}\right)}$, $I(X;Y;Z) = I(X;Y) - I(X;Y|Z)$

$I(X;Y|Z) = H(X|Z) - H(Y|Z) - H(X,Y|Z) = H(X|Z) - H(X|Y,Z)$

$$\cancel{I(X;Y) = \int_{-\infty}^{\infty}\int_{\infty}^{\infty} e^{\frac{-(x-\mu_x)^2}{2\sigma_x^2}} e^{\frac{-(y-\mu_y)^2}{2\sigma_y^2}} \ln \frac{e^{\frac{-(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2}}}{} \, dx\,dy} = I??? \text{ No!}$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

$$\cancel{H(X) = -\int_{-\infty}^{\infty} \frac{e^{\frac{-(x-\mu_x)^2}{2\sigma_x^2}}}{\sqrt{2\pi}\sigma_x} \ln\left(\frac{e^{\frac{-(x-\mu_x)^2}{2\sigma_x^2}}}{\sqrt{2\pi}\sigma_x}\right) dx = -\int_{-\infty}^{\infty} \frac{e^{\frac{-(x-\mu_x)^2}{2\sigma_x^2}}}{\sqrt{2\pi}\sigma_x} \cdot \left[\ln\left(\sqrt{2\pi}\sigma_x\right) - \left(\frac{(x-\mu_x)^2}{2\sigma_x^2}\right)\right] dx} \to \text{did this in first problem!}$$

$$\cancel{H(X) = \frac{1}{2}\left[1 + \ln(2\pi) + \ln(\sigma_x)\right] \to H(y) = \frac{1}{2}\left[1 + \ln(2\pi) + \ln(\sigma_y)\right]}$$

$$I(X;Y|Z) = \underbrace{H(X|Z)}_{P(x|z)} - \underbrace{H(X|Y,Z)}_{P(x|Y,z)} = \cancel{H(X|Z) - H(X|Y) - H(X|Y|Z)}?$$

$$H(X|Z) = -\int_{-\infty}^{\infty} \frac{e^{\frac{(x-\mu_x-\frac{\sigma_{xz}}{\sigma_z^2}(z-\mu_z))^2}{2(\sigma_x^2-\sigma_{xz}^2/\sigma_z^2)}}}{\sqrt{2\pi}\left(\sigma_x^2 - \frac{\sigma_{xz}^2}{\sigma_z^2}\right)^{1/2}} \ln(\dots)\,dx \to \text{did this for KL}$$

— but is this??

$$H(X|Z) = \frac{1}{2}\left[1 + \ln(2\pi) + \ln\left(\sigma_x^2 - \frac{\sigma_{xz}^2}{\sigma_z^2}\right)\right]$$

σ_{xy}

$$H(X|Y,Z) = -\int_{-\infty}^{\infty}\left(\sqrt{2\pi}\left(\sigma_x^2 - [\sigma_{xy}\ \sigma_{xz}]\begin{bmatrix} \sigma_y^2 & \sigma_{yz} \\ \sigma_{yz} & \sigma_z^2 \end{bmatrix}^{-1}\begin{bmatrix} \sigma_{xy} \\ \sigma_{xz} \end{bmatrix}\right)\right)^{-1} e^{\dots} \ln(\dots)\,dx$$

$$\to \text{KL cns} \to H(X|Y,Z) = \frac{[\sigma_{xy}\ \sigma_{xz}]}{\sigma_y^2\sigma_z^2 - \sigma_{yz}^2}\begin{bmatrix} \sigma_z^2 & -\sigma_{yz} \\ -\sigma_{yz} & \sigma_y^2 \end{bmatrix}\begin{bmatrix} \sigma_{xy} \\ \sigma_{xz} \end{bmatrix} = \frac{[\sigma_{xy}\ \sigma_{xz}]}{\sigma_y^2\sigma_z^2 - \sigma_{yz}^2}\begin{bmatrix} \sigma_z^2\sigma_{xy} - \sigma_{yz}\sigma_{xz} \\ \sigma_y^2\sigma_{xz} - \sigma_{yz}\sigma_{xy} \end{bmatrix}$$

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

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$$

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

$$I(X;Y|Z) = \frac{1}{2}\ln\left(\frac{\left(\sigma_x^2 - \frac{\sigma_{xz}^2}{\sigma_z^2}\right)\left(\sigma_y^2\sigma_z^2 - \sigma_{yz}^2\right)}{\sigma_x^2\sigma_y^2\sigma_z^2 - \sigma_x^2\sigma_{yz}^2 - \sigma_z^2\sigma_{xy}^2 - \sigma_y^2\sigma_{xz}^2 + 2\sigma_{xz}\sigma_{yz}\sigma_{xy}}\right)$$

xz, yz, xy

7) (Linear Regression) $J(\underline{w}) = \frac{1}{N}\sum_{n=1}^{N}(z_n - \underline{w}^T \underline{x}_n)^2 = \frac{1}{N}\|\underline{z} - X\underline{w}\|_2^2$ $\quad X = \begin{bmatrix} \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \end{bmatrix}$ $N > P$

$\underline{z} = X\underline{w} + \underline{v}$

a) $\nabla_{\underline{w}} J = -\frac{2}{N}\sum_{n=1}^{N}(z_n - \underline{w}^T \cancel{x_n})x_n \Leftrightarrow \nabla_{\underline{w}} J = \frac{2X^T}{N}(\underline{z} - X\underline{w}) = 0 = X^T\underline{z} = X^T X\underline{w}$

$\rightarrow$ ~~...~~ $\underline{w} = (X^{-1})(X^T)^{-1}X^T\underline{z} = \boxed{(X^T X)^{-1}X^T\underline{z} = \hat{\underline{w}}}$

b) $\hat{\underline{z}} = X\hat{\underline{w}} = X(X^T X)^{-1}X^T z = X(X^T X)^{-1}X^T(\cancel{\ldots})(X\underline{w} + \underline{v})$

$= X(X^T X)^{-1}(X^T X)\underline{w} + X(X^T X)^{-1}X^T\underline{v}$

$= \boxed{X\underline{w} + X(X^T X)^{-1}X^T\underline{v} = X\underline{w} + P\underline{v} = \hat{\underline{z}}}$

c) $\underline{z} - \hat{\underline{z}} = \cancel{\underline{z} - X\hat{\underline{w}} = \ldots} \underbrace{\cancel{\underline{v}}}_{=I\underline{v}} \cancel{- X\underline{w} - X(X^T X)^{-1}X^T\underline{v}} = X\underline{w} - X\underline{w} + \underline{v} - P\underline{v}$

$\cancel{= (I-P)\underline{v}}$

$= I\underline{v} - P\underline{v} = \boxed{\begin{matrix}(I-P)\underline{v} \\ = z - \hat{z}\end{matrix}}$

d) $E[J(\underline{w})|_{\underline{w}=\hat{\underline{w}}}] = E[\frac{1}{N}\|\underline{z} - X\hat{\underline{w}}\|_2^2] = \frac{1}{N}E[\|\underline{z} - \hat{\underline{z}}\|_2^2]$

$= \frac{1}{N}E[\|(I-P)\underline{v}\|_2^2] = \frac{1}{N}E[((I-P)\underline{v})^T(I-P)\underline{v}] = \frac{1}{N}E[\underline{v}^T \underbrace{(I-P)^2}_{\text{constant}}\underline{v}]$

$= \frac{1}{N}\underbrace{E[\underline{v}^T\underline{v}]}_{\int \frac{x^2 e^{\ldots/\sigma_2}}{\sigma_2} = \sigma^2}\, tr[(I-P)]^2 = \frac{\sigma^2}{N}(N-P) = \boxed{\sigma^2(1 - P/N) = E[J(\underline{w})|_{\underline{w}=\hat{\underline{w}}}]}$

$\hookrightarrow tr[(I-P)^2] = tr[(I-P)^T(I-P)] = tr(I-P) = N - P$

$X(X^T X)^{-1}X^T$ $\quad N\times P \;\; P\times P \;\; P\times N : N\times N$ $\quad \cancel{\ldots}$?

$\sim/26$

e) $\cancel{x_{N+1}}\; z_{N+1} - \hat{z}_{N+1} = z_{N+1} - \underline{x}_{N+1}^T\hat{\underline{w}} = \cancel{z_{N+1} - \underline{x}_{N+1}^T(X^T X)^{-1}X^T z}$

$= \underline{x}_{N+1}^T\underline{w} + v_{N+1} - \underline{x}_{N+1}^T\hat{\underline{w}} = \underline{x}_{N+1}^T\underline{w} + v_{N+1} - \underline{x}_{N+1}^T(X^T X)^{-1}X^T(X\underline{w} + \underline{v}_{\ldots})$

$= \underline{x}_{N+1}^T\underline{w} + v_{N+1} - \underline{x}_{N+1}^T\underbrace{(X^T X)^{-1}(X^T X)}_{I}\underline{w} - \underline{x}_{N+1}^T \cancel{\ldots}(X^T X)^{-1}X^T\underline{v}_{\ldots}$

$= \boxed{\cancel{x_{N+1} - x_{N+1}}\, v_{N+1} = z_{N+1} - \hat{z}_{N+1}} = \boxed{v_{N+1} - \underline{x}_{N+1}^T(X^T X)^{-1}X^T\underline{v}_{\ldots} = z_{N+1} - \hat{z}_{N+1}}$

f) $E[(z_{N+1} - \hat{z}_{N+1})^2] = E[(v_{N+1} - x_{N+1}^T(X^T X)^{-1}X^T\underline{v})^2] = E[v_{N+1}^2 - 2v_{N+1}x_{N+1}^T(X^T X)^{-1}\underline{v} + x_{N+1}^T(X^T X)^{-1}X^T\underline{v}\,\underline{v}^T X(X^T X)^{-1}x_{N+1}]$

$\leftarrow AA^T = A^T A$

$= E[v_{N+1}^2] - 2E[x_{N+1}^T(X^T X)^{-1}X^T\underline{v}] + E[x_{N+1}^T(X^T X)^{-1}X^T\underbrace{\underline{v}\,\underline{v}^T}_{\text{scalar}}X(X^T X)^{-1}x_{N+1}]$

$= \sigma^2 \cancel{- 2\sigma\, tr[x_{N+1}(X^T X)^{-1}X^T]} - 2\cdot 0 + tr[\cancel{x_{N+1}}(X^T X)^{-1}\underbrace{X^T X(X^T X)^{-1}}_{I}\overbrace{E[\underline{v}\,\underline{v}^T]}^{\sigma^2}\overbrace{E[x_{N+1}^T x_{N+1}]}^{\Sigma/N-1}]$?

$= \sigma^2 + \frac{\sigma^2}{N-1}\underbrace{tr[(X^T X)^{-1}]}_{\Sigma^{-1}}\cancel{\ldots}\Sigma = \sigma^2 + \frac{\sigma^2}{N-1}\underbrace{tr[\Sigma^{-1}\Sigma]}_{P} = \sigma^2(1 + \frac{P}{N-1}) \approx \boxed{\sigma^2(1 + \frac{P}{N}) \approx E[(z_{N+1} - \hat{z}_{N+1})^2]}$

g) $\hat{z}_{-i} = \underline{\hat{w}}_{-i}^T \underline{x}_i = \left(\underbrace{(X^TX)^{-1}X^T\underline{z}_i}\right)^T \underline{x}_i = \underline{z}_i^T \underbrace{X(X^TX)^{-1}\underline{x}_i}_{\hookrightarrow\, ii^{th} \text{ element of } P} =$

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

$\hat{z}_i = X\hat{\underline{w}} = X\underline{w} + P\underline{v} = \underline{w}_i^T x_i + v$

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

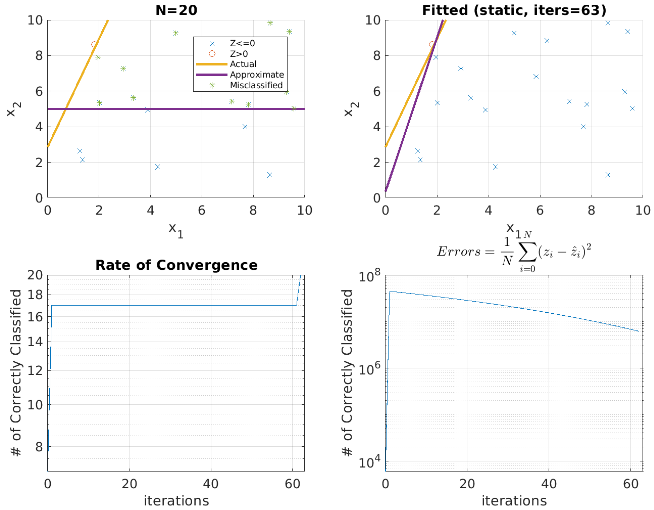
i) $E[J_{LOOCV}] = \frac{1}{N}\sum_{i=1}^{N} \dfrac{E[(z_i - \hat{z}_i)^2]}{(1-P_{ii})^2}$ ← use part (d) of $\Rightarrow E[J(\underline{\hat{w}})] = \sigma^2(1 - P/N)$

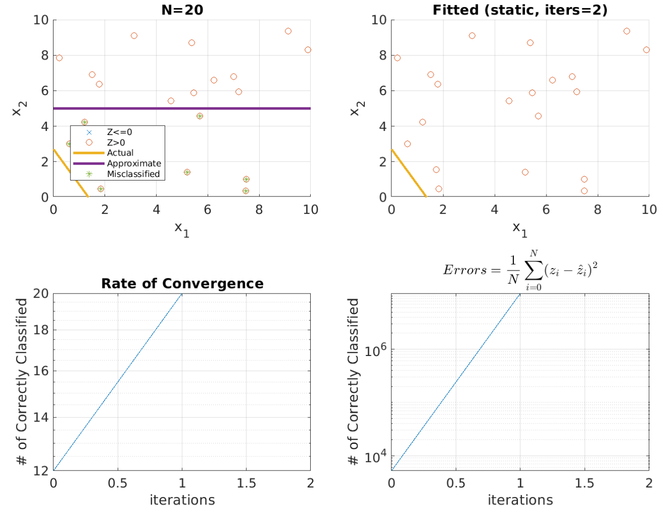
except $tr[I] = 1$ & use the ith component of P instead of $tr[P] \Rightarrow tr[P] = P_{ii}$ → because it's discrete

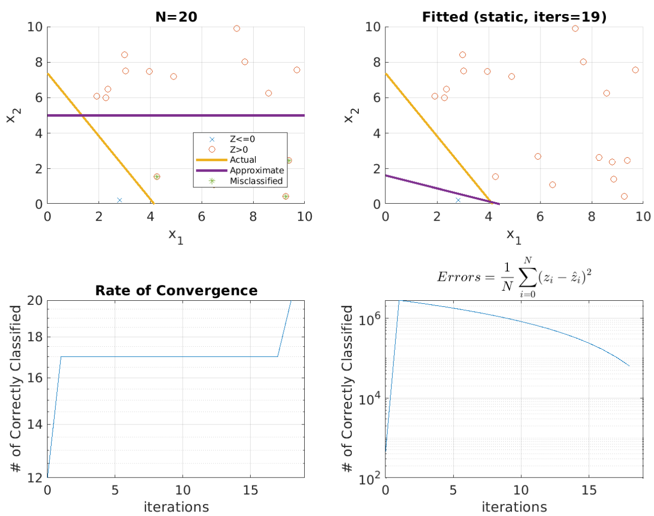
$= \boxed{\dfrac{\sigma^2}{N}\sum \dfrac{1+P_{ii}}{(1-P_{ii})^2} = E[J_{LOOCV}]}$

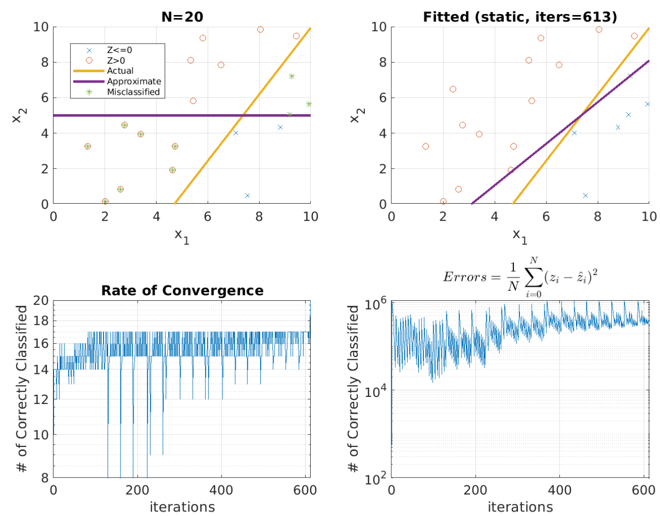
The "static" method does well at low N which makes sense since you are only dealing w/ a few data points so you can attack them one at a time.

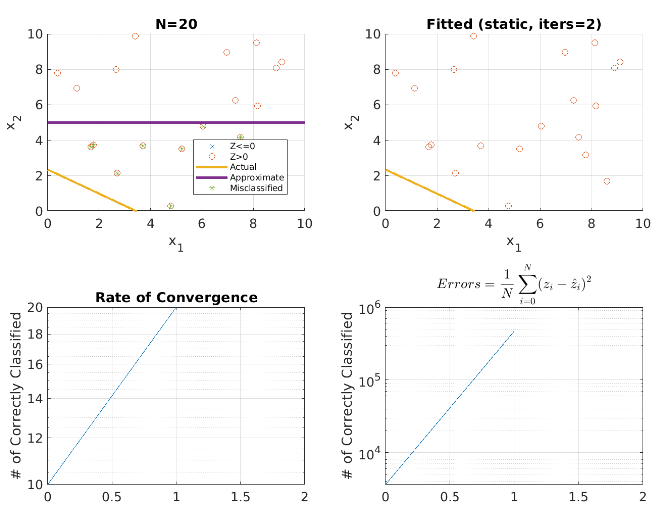
The "shuffled" method does well at large N since you are dealing with more of an average of data points rather than one at a time. This helps to avoid solving the regression for only one or two data points.

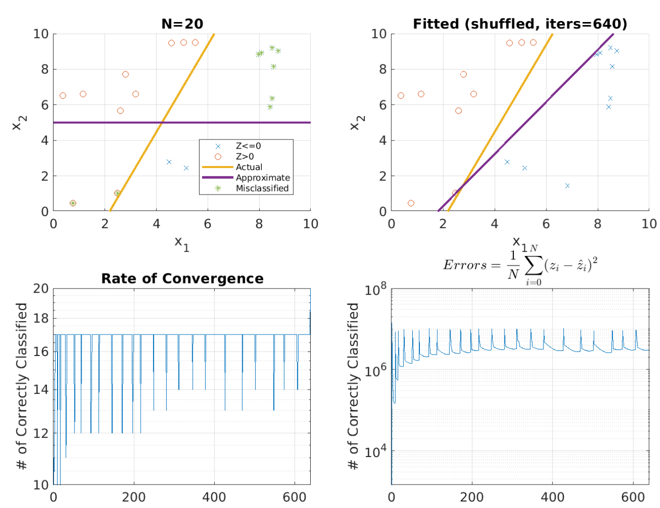

N=20
Fitted (static, iters=63)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

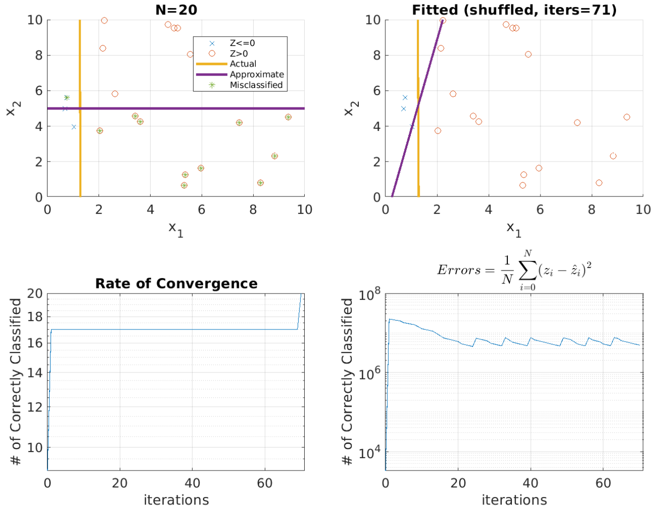

N=20
Fitted (static, iters=2)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

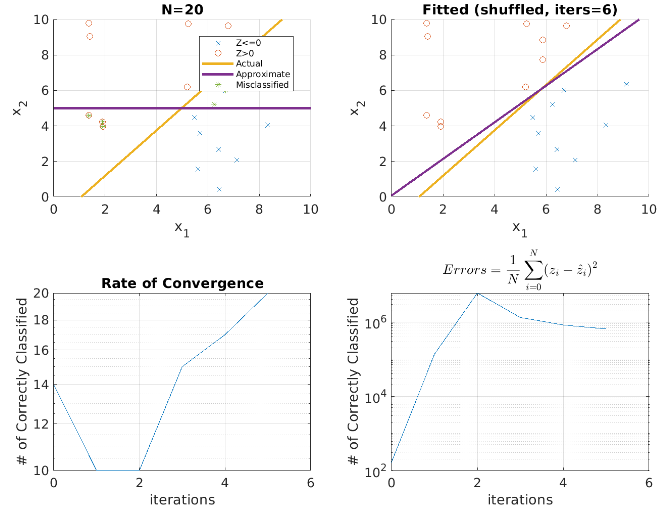

N=20
Fitted (static, iters=19)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

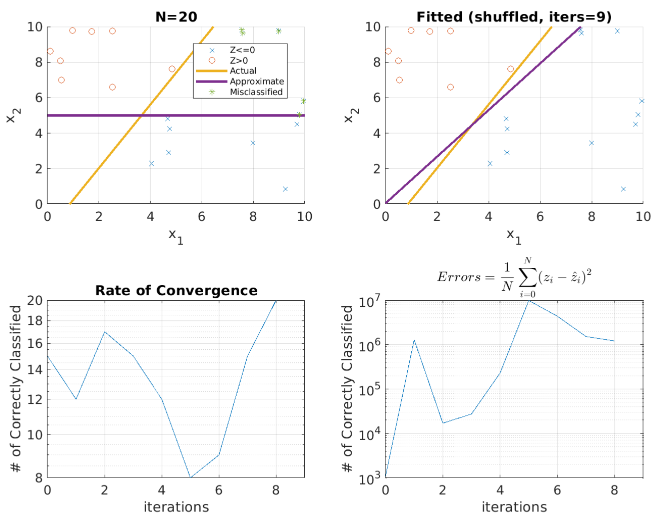

N=20
Fitted (static, iters=613)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

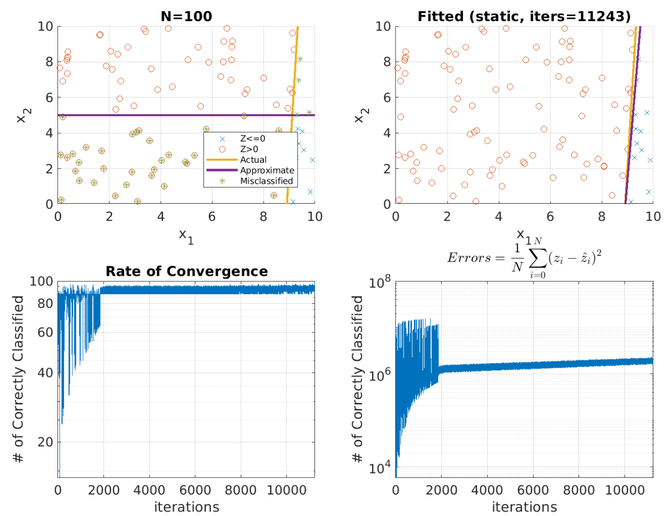

N=20
Fitted (static, iters=2)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

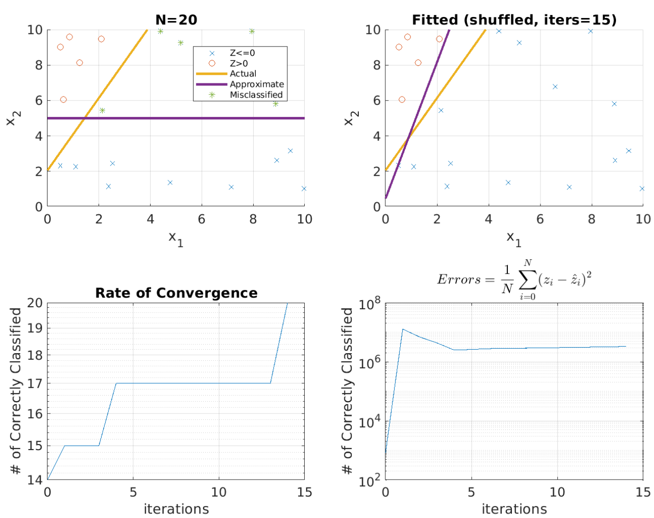

N=20
Fitted (shuffled, iters=640)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

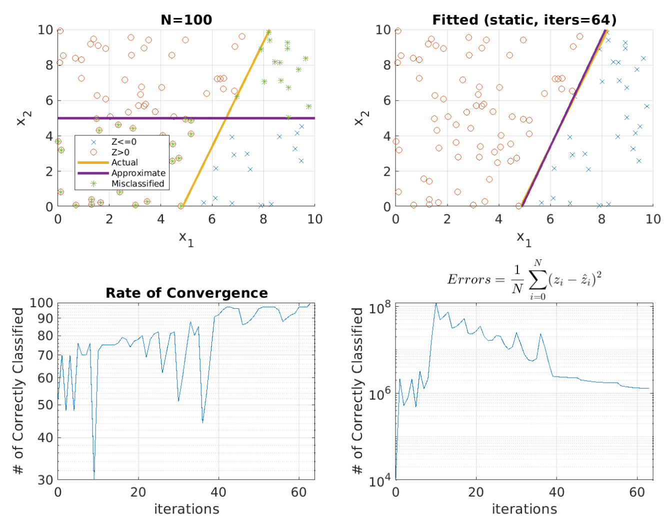

N=20
Z<=0
Z>0
Actual
Approximate
Misclassified
Fitted (shuffled, iters=71)
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

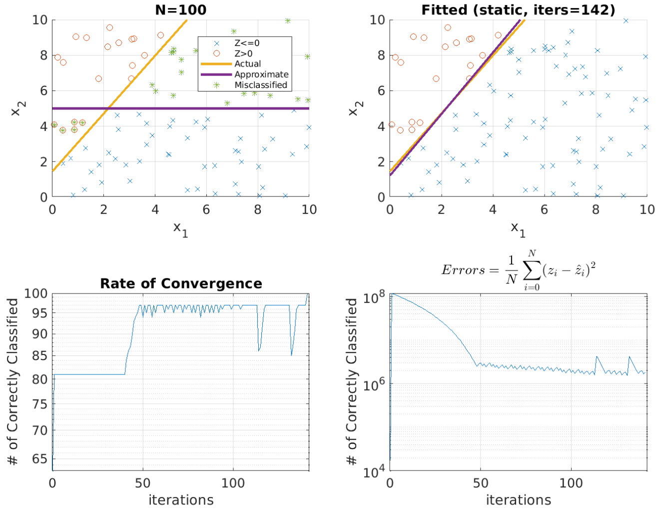

N=20
Z<=0
Z>0
Actual
Approximate
Misclassified
Fitted (shuffled, iters=6)
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

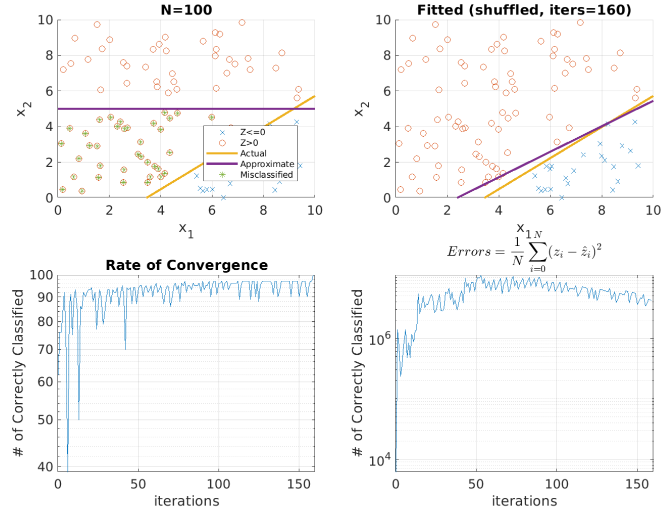

N=20
Z<=0
Z>0
Actual
Approximate
Misclassified
Fitted (shuffled, iters=9)
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

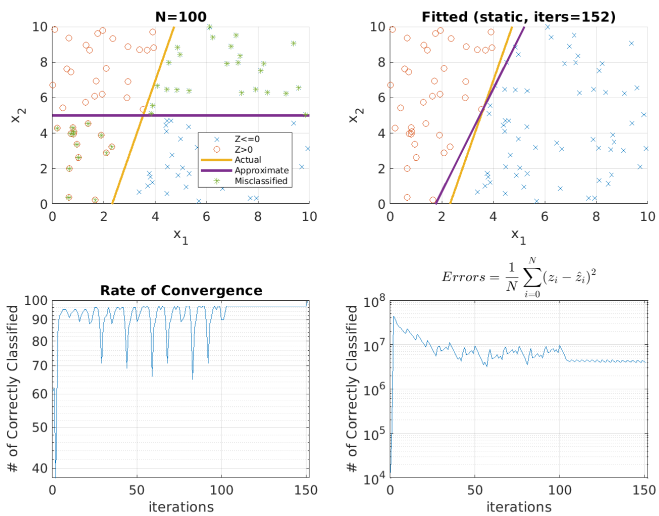

N=100
Z<=0
Z>0
Actual
Approximate
Misclassified
Fitted (static, iters=11243)
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

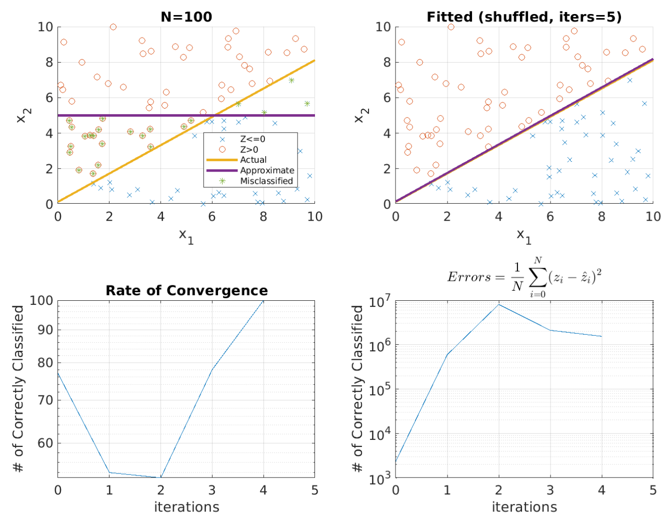

N=20
Z<=0
Z>0
Actual
Approximate
Misclassified
Fitted (shuffled, iters=15)
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

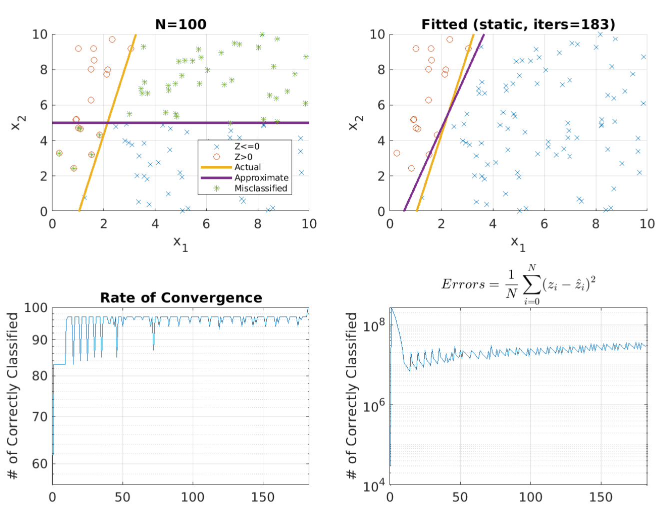

N=100
Z<=0
Z>0
Actual
Approximate
Misclassified
Fitted (static, iters=64)
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

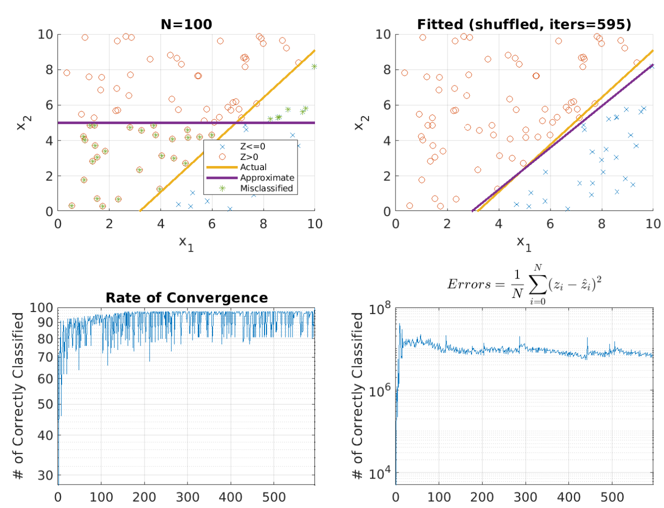

N=100
Fitted (static, iters=142)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

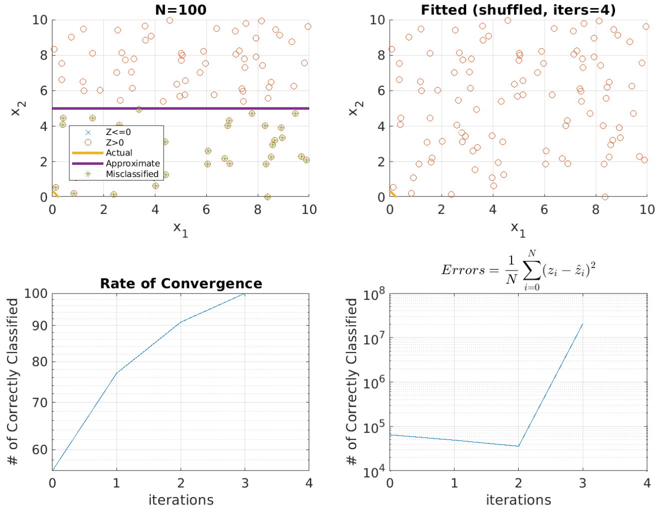

N=100
Fitted (shuffled, iters=160)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

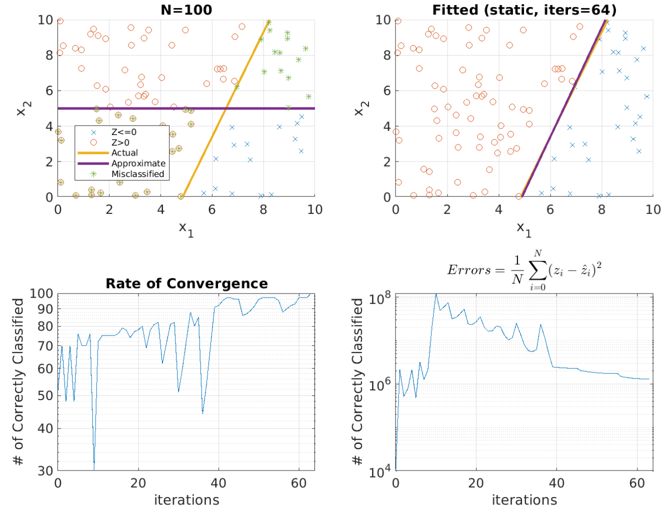

N=100
Fitted (static, iters=152)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

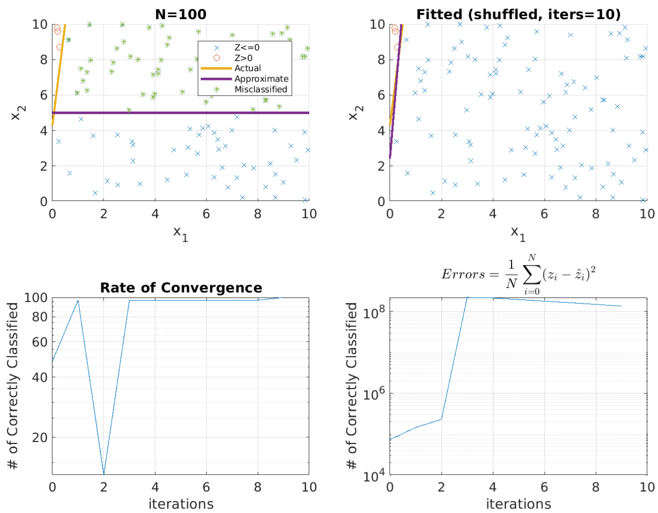

N=100
Fitted (shuffled, iters=5)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

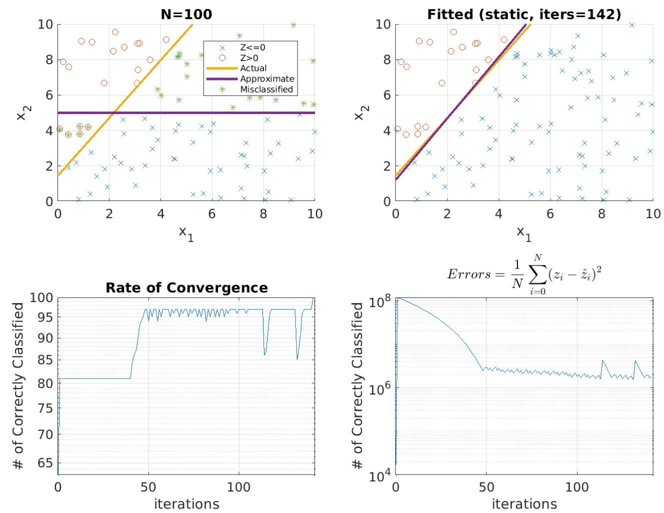

N=100
Fitted (static, iters=183)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

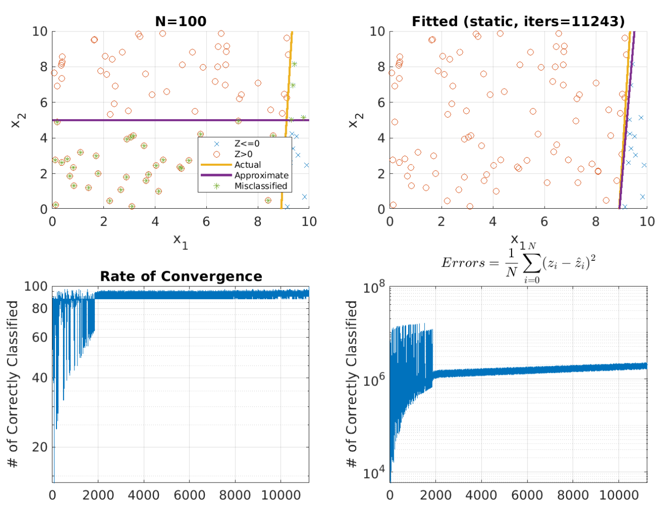

N=100
Fitted (shuffled, iters=595)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

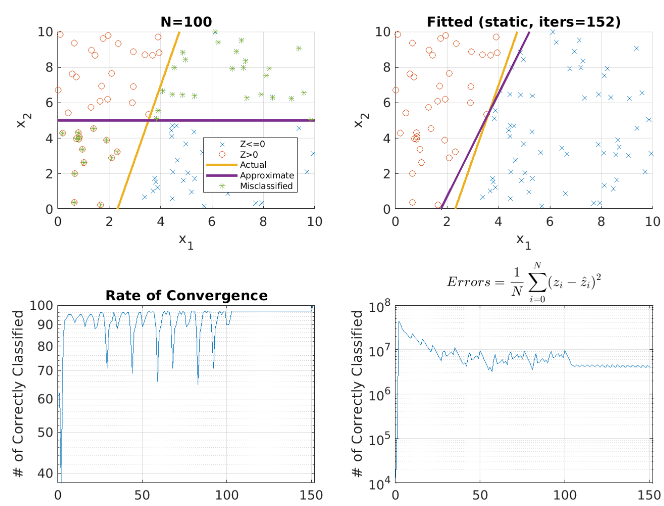

N=100
Fitted (shuffled, iters=4)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations
x_1
x_2

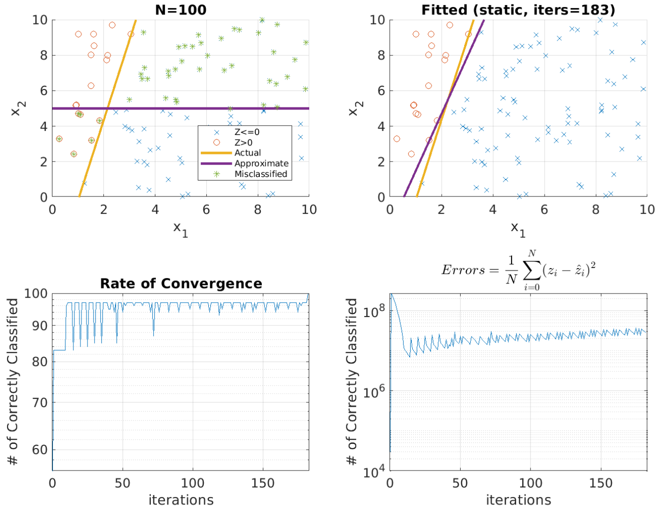

N=100
Fitted (static, iters=64)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations
x_1
x_2

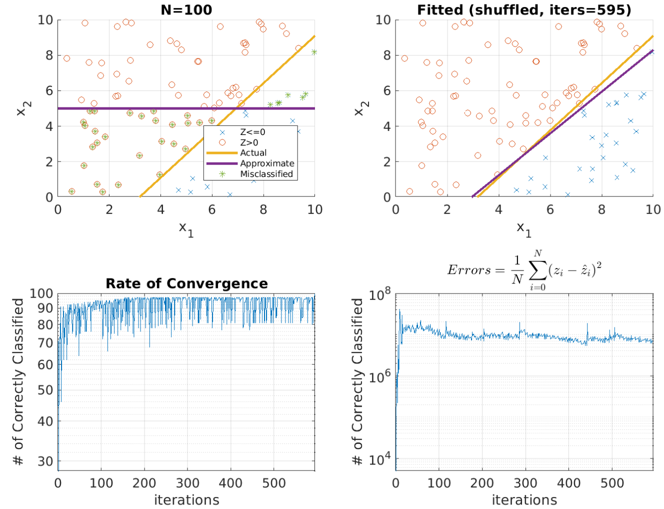

N=100
Fitted (shuffled, iters=10)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations
x_1
x_2

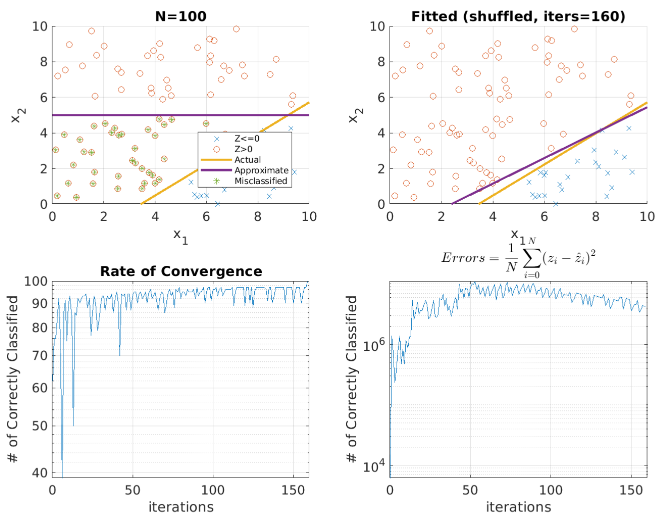

N=100
Fitted (static, iters=142)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations
x_1
x_2

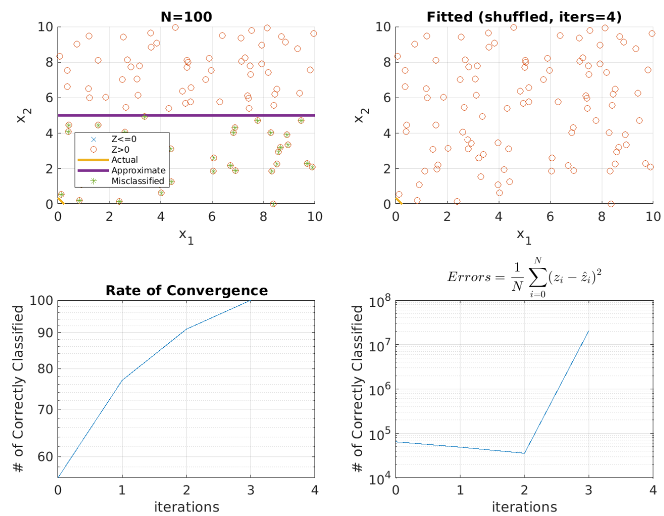

N=100
Fitted (static, iters=11243)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations
x_1
x_2

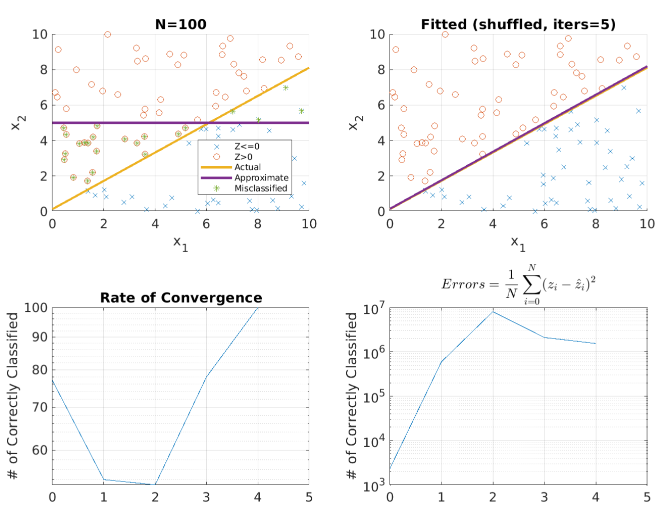

N=100
Fitted (static, iters=152)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations
x_1
x_2

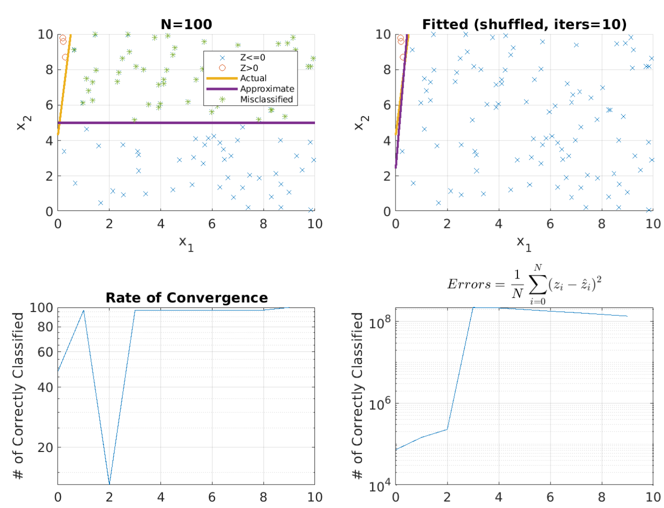

N=100
Fitted (static, iters=183)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

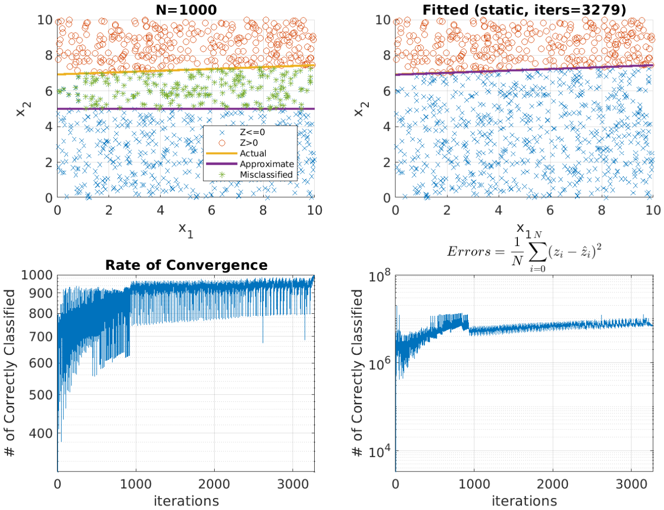

N=100
Fitted (shuffled, iters=595)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

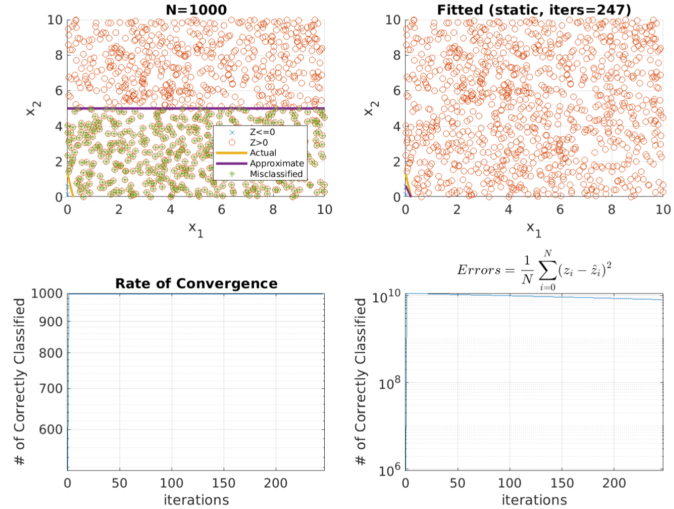

N=100
Fitted (shuffled, iters=160)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

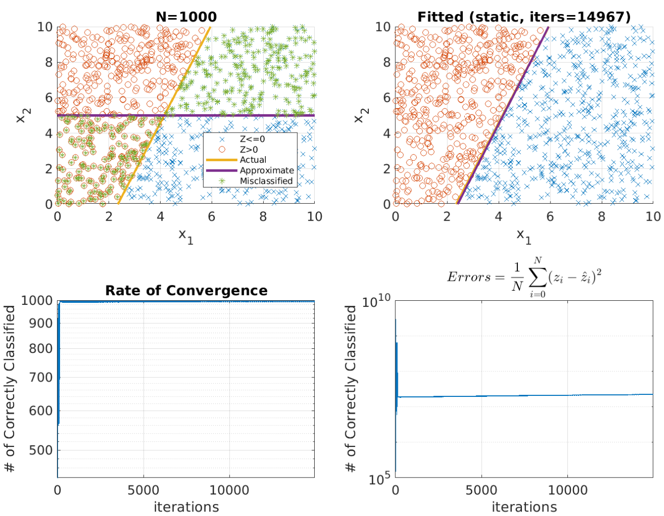

N=100
Fitted (shuffled, iters=4)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

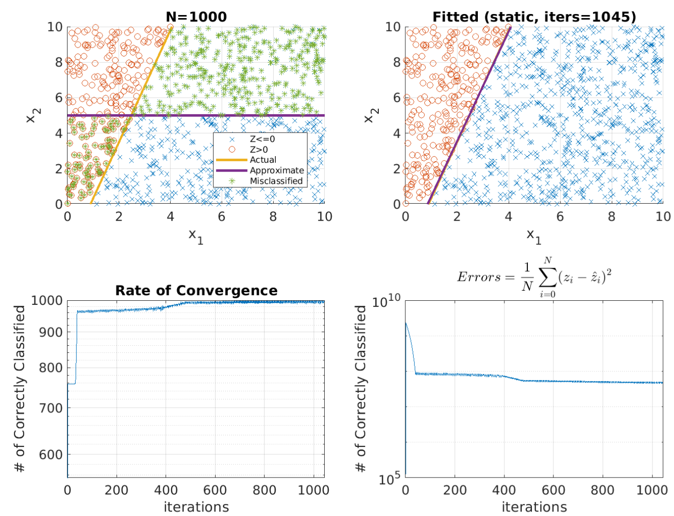

N=100
Fitted (shuffled, iters=5)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

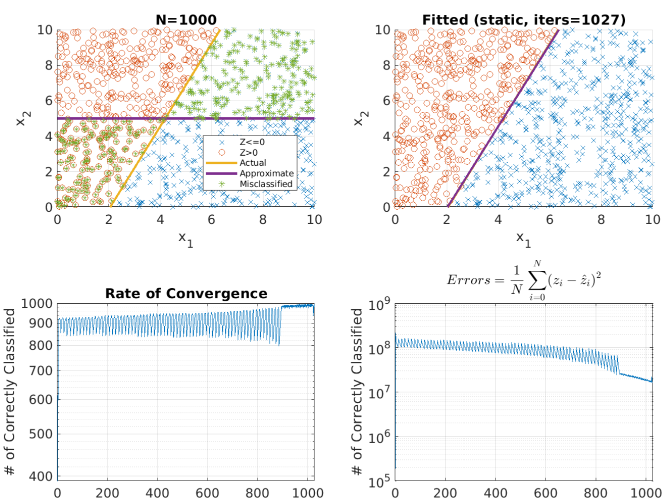

N=100
Fitted (shuffled, iters=10)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations

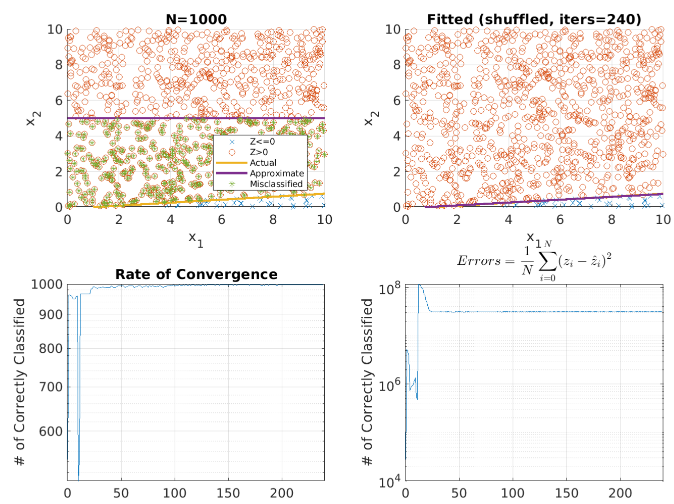

N=1000
Fitted (static, iters=3279)
Z<=0
Z>0
Actual
Approximate
Misclassified
x_1
x_2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

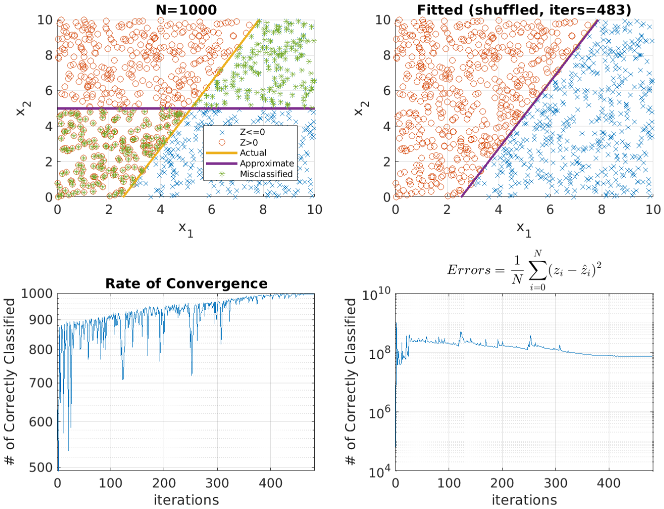

N=1000
Fitted (static, iters=247)
Z<=0
Z>0
Actual
Approximate
Misclassified
x_1
x_2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$


N=1000
Fitted (static, iters=14967)
Z<=0
Z>0
Actual
Approximate
Misclassified
x_1
x_2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

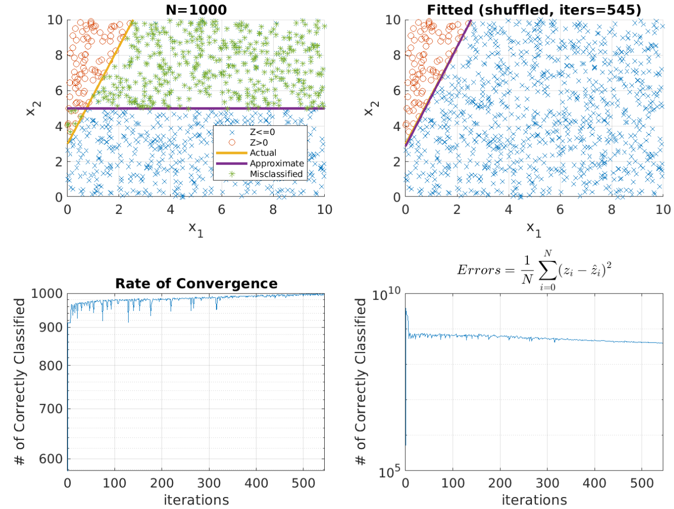

N=1000
Fitted (static, iters=1045)
Z<=0
Z>0
Actual
Approximate
Misclassified
x_1
x_2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$


N=1000
Fitted (static, iters=1027)
Z<=0
Z>0
Actual
Approximate
Misclassified
x_1
x_2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

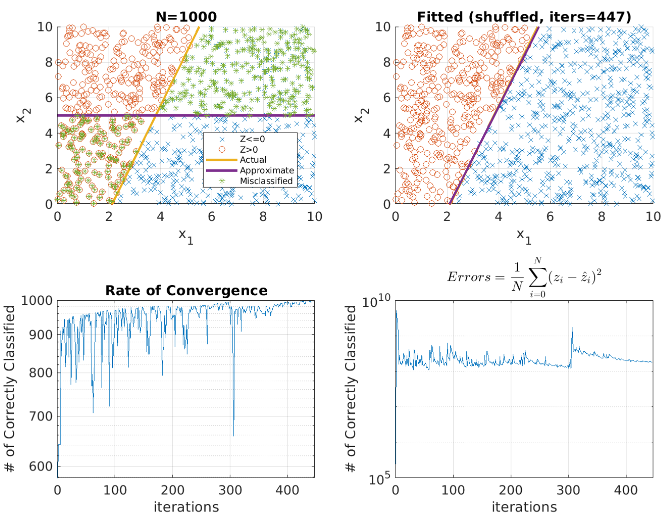

N=1000
Fitted (shuffled, iters=240)
Z<=0
Z>0
Actual
Approximate
Misclassified
x_1
x_2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$



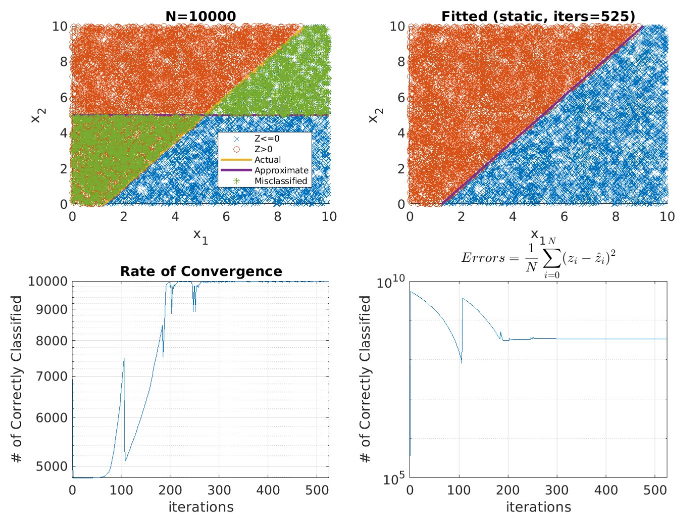
N=1000
Fitted (shuffled, iters=483)
Z<=0
Z>0
Actual
Approximate
Misclassified
x_1
x_2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$



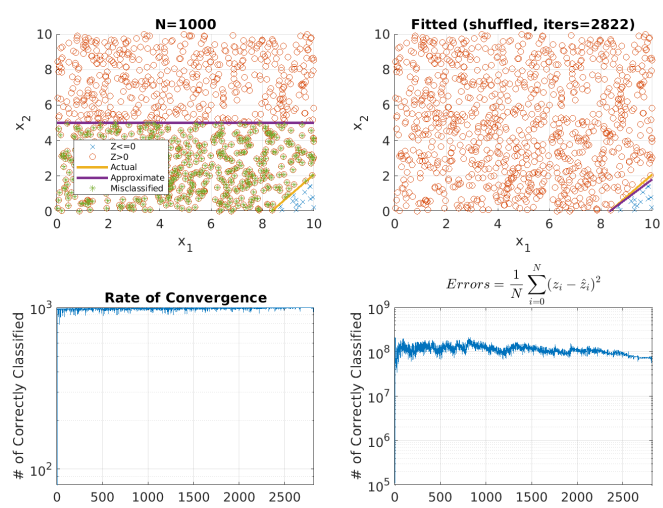
N=1000
Fitted (shuffled, iters=545)
Z<=0
Z>0
Actual
Approximate
Misclassified
x_1
x_2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$



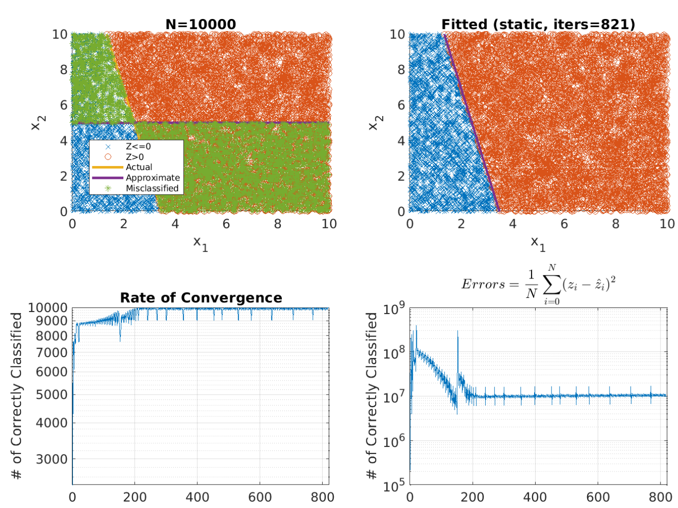
N=1000
Fitted (shuffled, iters=447)
Z<=0
Z>0
Actual
Approximate
Misclassified
x_1
x_2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$



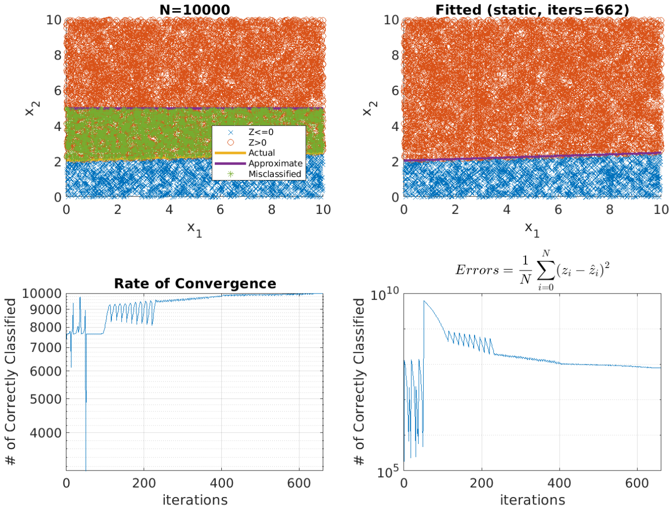
N=10000
Fitted (static, iters=525)
Z<=0
Z>0
Actual
Approximate
Misclassified
x_1
x_2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

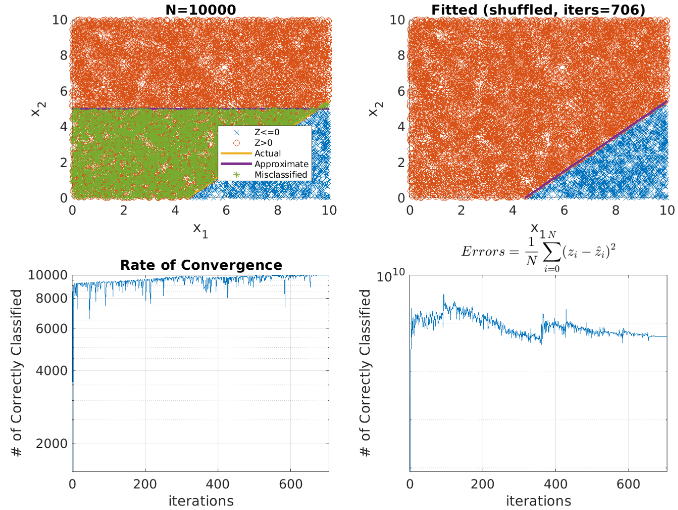


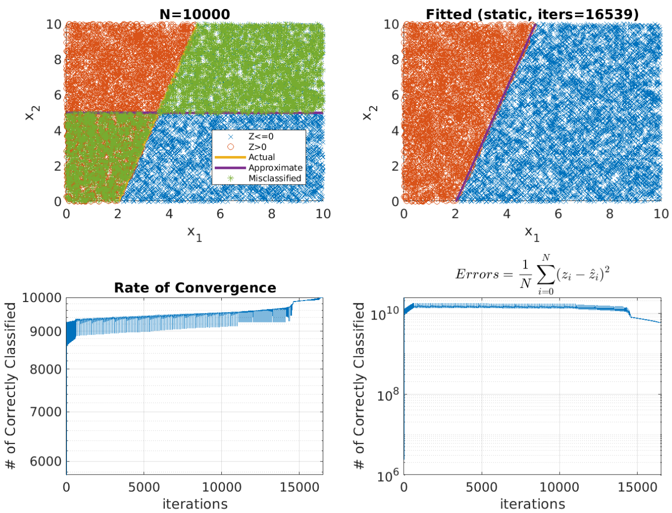
N=1000
Fitted (shuffled, iters=2822)
Z<=0
Z>0
Actual
Approximate
Misclassified
x_1
x_2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

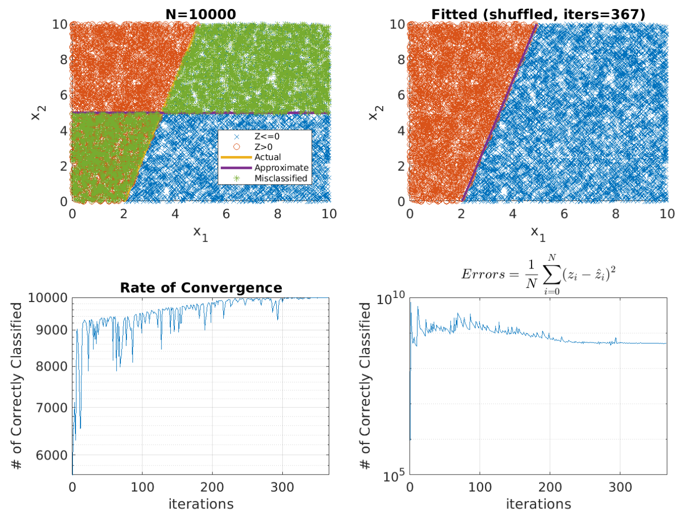


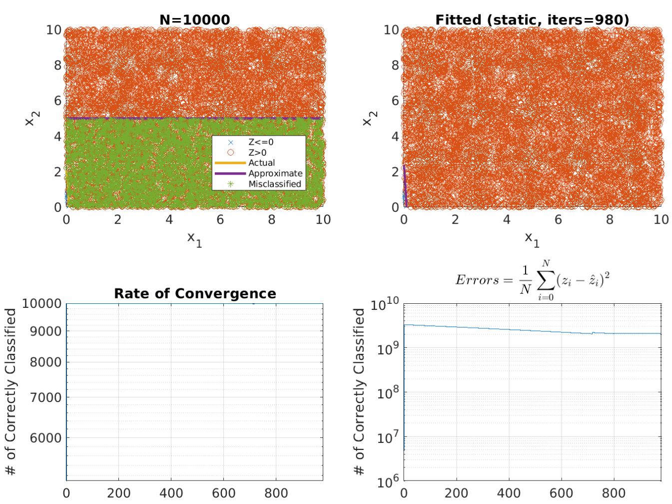
N=10000
Fitted (static, iters=821)
Z<=0
Z>0
Actual
Approximate
Misclassified
x_1
x_2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

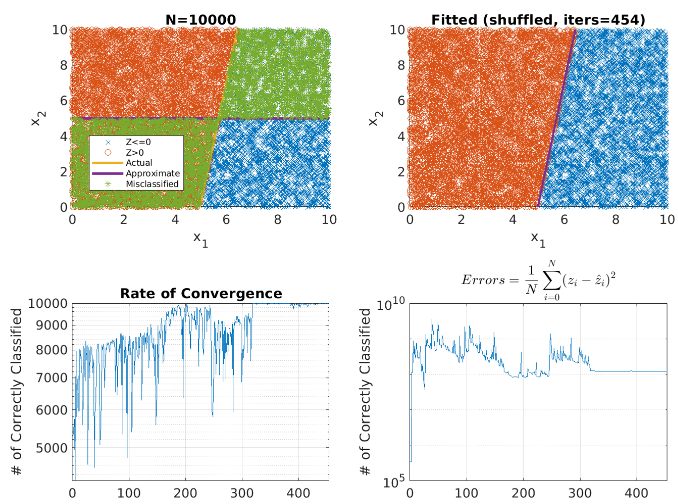

N=10000
Fitted (static, iters=662)
Z<=0
Z>0
Actual
Approximate
Misclassified
x1
x2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

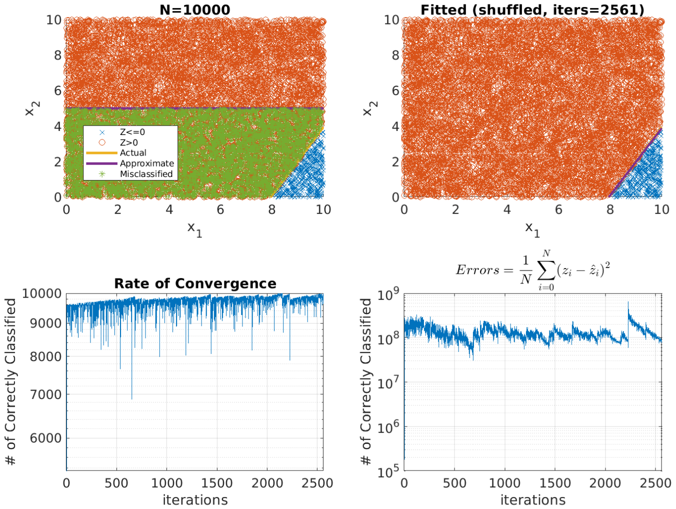

N=10000
Fitted (shuffled, iters=706)
Z<=0
Z>0
Actual
Approximate
Misclassified
x1
x2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

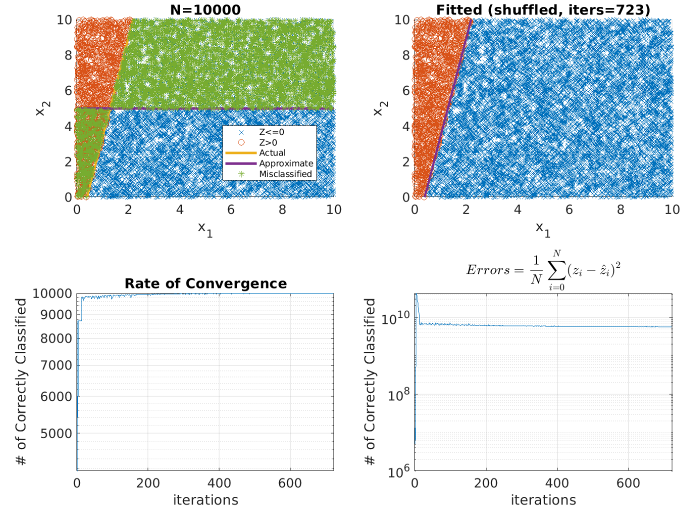

N=10000
Fitted (static, iters=16539)
Z<=0
Z>0
Actual
Approximate
Misclassified
x1
x2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

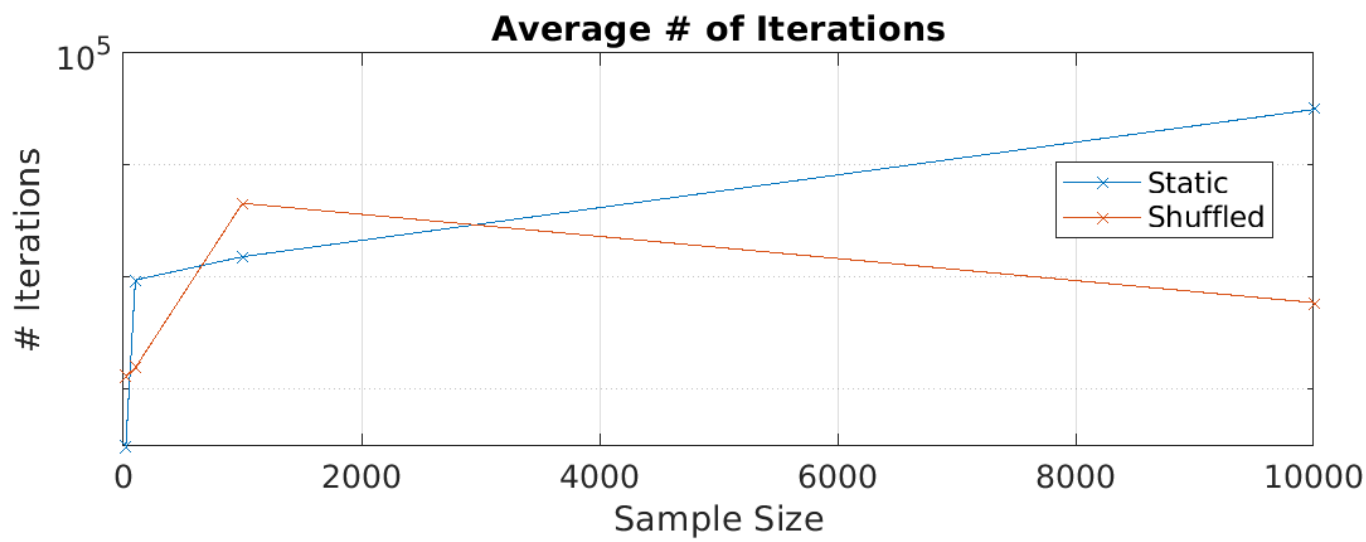

N=10000
Fitted (shuffled, iters=367)
Z<=0
Z>0
Actual
Approximate
Misclassified
x1
x2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$

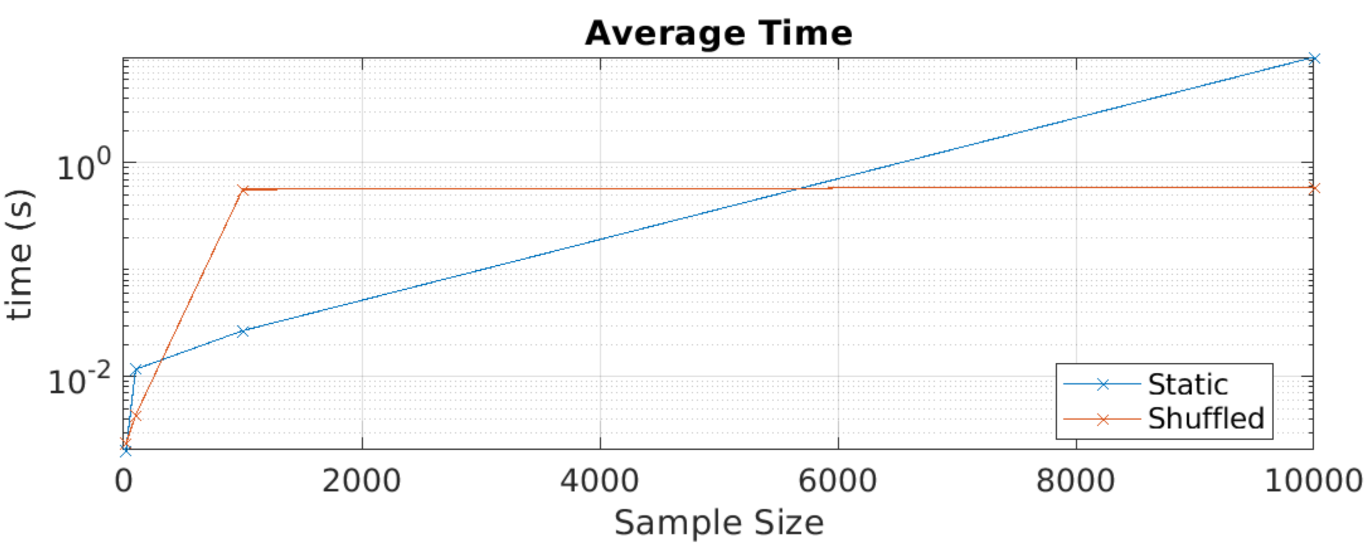

N=10000
Fitted (static, iters=980)
Z<=0
Z>0
Actual
Approximate
Misclassified
x1
x2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$


N=10000
Fitted (shuffled, iters=454)
Z<=0
Z>0
Actual
Approximate
Misclassified
x1
x2
Rate of Convergence
# of Correctly Classified
iterations
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$


N=10000
Fitted (shuffled, iters=2561)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations


N=10000
Fitted (shuffled, iters=723)
Z<=0
Z>0
Actual
Approximate
Misclassified
Rate of Convergence
$Errors = \frac{1}{N}\sum_{i=0}^{N}(z_i - \hat{z}_i)^2$
# of Correctly Classified
iterations


Average # of Iterations
Iterations
Sample Size
Static
Shuffled


Average Time
time (s)
Sample Size
Static
Shuffled