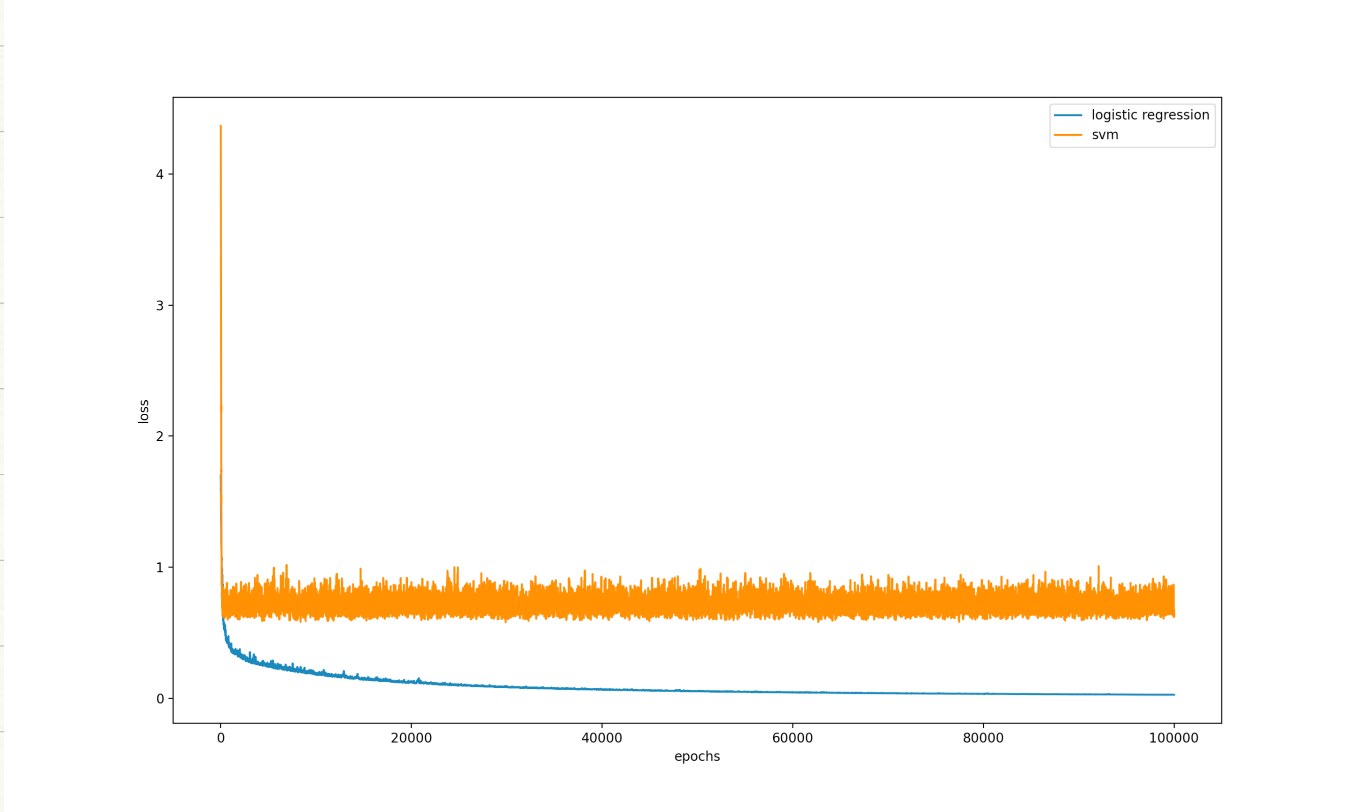
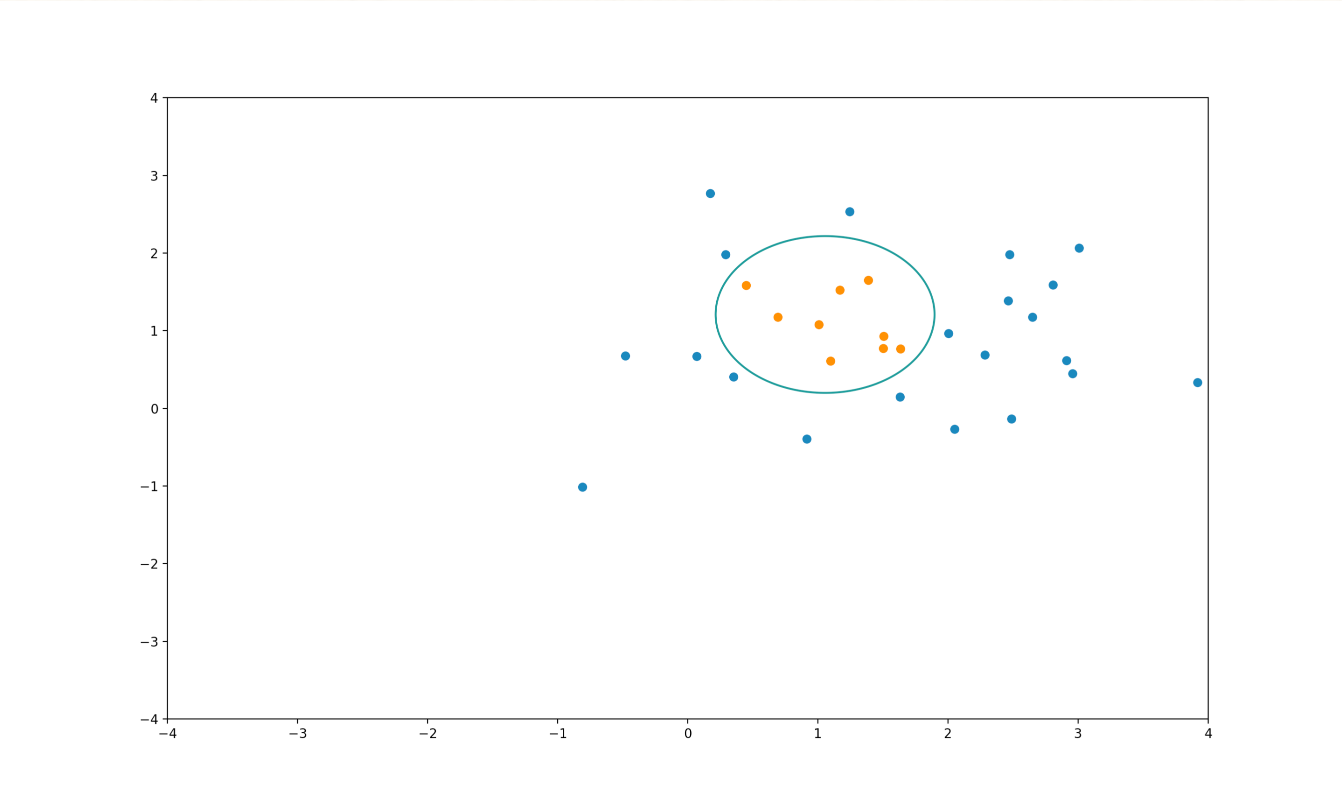
P1&P2



P3



P4

Proposition) $-\log x$ is a strongly convex function. (*)

pf)

$f(x): \mathbb{R}^+ \to \mathbb{R} \triangleq -\log x$, WLOG, choose $x_2 < x_1$ from $\mathbb{R}^+$

Since $x_1, x_2 \Rightarrow \eta x_1 + (1-\eta)x_2 \in \mathbb{R}^+$, $\forall \eta \in (0,1)$, $\mathbb{R}^+$ is convex.

$f''(x) = \frac{1}{x^2} > 0 \Rightarrow f'(x)$ is strictly increasing function $\cdots$ (1)

$g(t) \triangleq t f(x_1) + (1-t) f(x_2) - f(tx_1 + (1-t)x_2)$, $\quad g: [0,1] \to \mathbb{R}$

$g(t) = t(f(x_1) - f(tx_1+(1-t)x_2)) - (1-t)(f(tx_1+(1-t)x_2) - f(x_2))$

$= t(1-t)(x_1-x_2) f'(c_1) - (1-t)t(x_1-x_2) f'(c_2)$, $\exists!\, c_1 \in (tx_1+(1-t)x_2, x_1)$, $\exists!\, c_2 \in (x_2, tx_1+(1-t)x_2)$, by MVT & (1)

$= t(1-t)(x_1-x_2)(f'(c_1) - f'(c_2)) = t(1-t)(x_1-x_2)(c_1-c_2) f''(c_3)$, $\exists!\, c_3 \in (c_2, c_1)$ by MVT & (1)

$\therefore g(t) > 0$ for $t \in (0,1)$. ※

$D_{KL} = \sum_{i=1}^n p_i \left(\log \frac{p_i}{q_i}\right)$, $\sum_{i=1}^n p_i = 1$, $\sum_{i=1}^n q_i = 1$, $\{p\}_i, \{q\}_i \geq 0$

$= \sum_{i=1}^n p_i \left(-\log \frac{q_i}{p_i}\right)$

$\varphi(x) \triangleq -\log x$, X is r.v. s.t. $f(X = \frac{q_i}{p_i}) = p_i$, $\forall i$

Then, $D_{KL} = E[\varphi(X)] \geq \varphi(E[X]) = \varphi\left(\sum_{i=1}^n p_i \cdot \frac{q_i}{p_i}\right) = \varphi(1) = 0$ ※

P5.

Let) X be r.v. s.t. $f(X=\frac{q_i}{p_i}) = p_i$, $\forall i$. Since, $p \neq q$, X isn't constant r.v.

$\varphi(x) \triangleq -\log x$ is strictly convex function. ASW, $D_{KL} = E[\varphi(X)] > \varphi(E[X]) = 0$ ※

P6

$$f_\theta(x) = u^T \sigma(ax+b) = \sum_{j=1}^{p} u_j \sigma(a_j x + b_j)$$

$a, b, u \in \mathbb{R}^p,\quad \theta \in \mathbb{R}^{3p}$

i) $\nabla_u f_\theta(x) = \sigma(ax+b)$

$$\frac{\partial}{\partial u_i} f_\theta = \frac{\partial}{\partial u_i}\left(\sum_{j=1}^{p} u_j \sigma(a_j x + b_j)\right) = \sigma(a_i x + b_i)$$

$\therefore \nabla_u f_\theta(x) = \sigma(ax+b)$

ii) $\nabla_b f_\theta(x) \overset{①}{=} \sigma'(ax+b) \odot u \overset{②}{=} \mathrm{diag}(\sigma'(ax+b))\, u$

① $\frac{\partial}{\partial b_i} f_\theta = \frac{\partial}{\partial b_i}\left(\sum_{j=1}^{p} u_j \sigma(a_j x + b_j)\right) = u_i \sigma'(a_i x + b_i)$

$\sigma'(ax+b) \in \mathbb{R}^p,\quad u \in \mathbb{R}^p \Rightarrow \odot$ is well-defined.

$(\sigma'(ax+b))_i = \sigma'(ax_i + b_i), \quad (u)_i = u_i$

$(\sigma'(ax+b) \odot u)_i = \sigma'(ax_i + b)u_i = \frac{\partial f_\theta}{\partial b_i} = (f_\theta(x))_i$

② $\mathrm{diag}(\sigma'(ax+b))\, u = \sum_{i=1}^{p} \mathrm{diag}(\sigma'(ax+b))_{:,i}\, u_i \in \mathbb{R}^p$

$(\mathrm{diag}(\sigma'(ax+b)))_i = \left(\sum_{j=1}^{p} \mathrm{diag}(\sigma'(ax+b))_{:,j}\, u_j\right)_i = \sigma'(ax_i + b)u_i = (\sigma'(ax+b) \odot u)_i$

$\therefore \nabla_b f_\theta(x) = \sigma'(ax+b) \odot u = \mathrm{diag}(\sigma'(ax+b))\, u$

iii) $\nabla_a f_\theta(x) \overset{①}{=} (\sigma'(ax+b) \odot u)\, x \overset{②}{=} \mathrm{diag}(\sigma'(ax+b))\, ux$

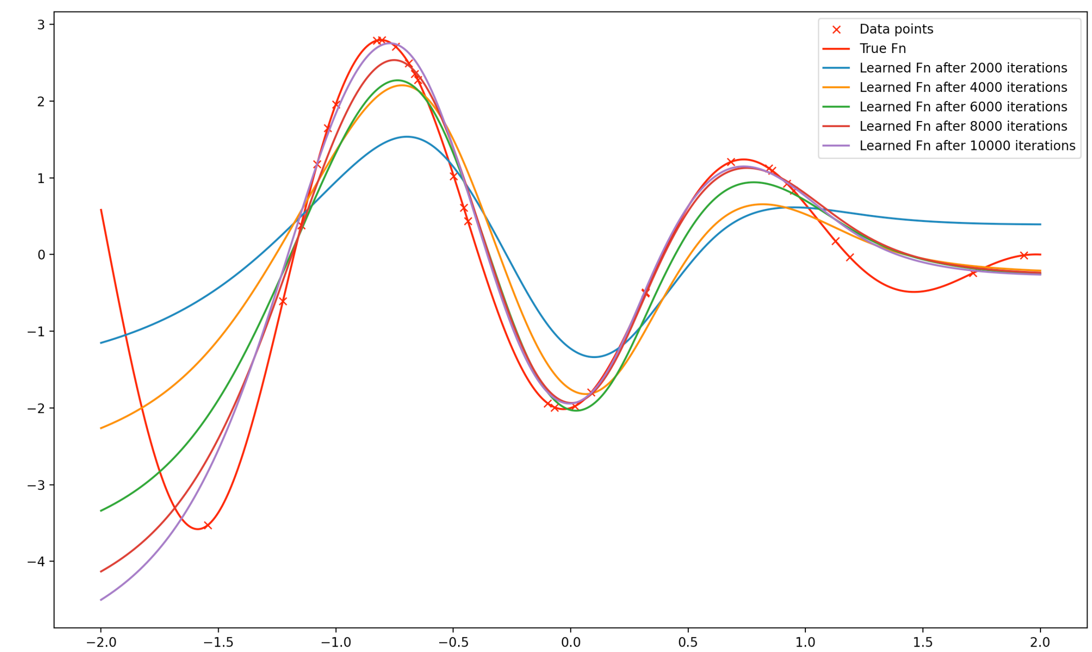
① $\frac{\partial}{\partial a_i} f_\theta = \frac{\partial}{\partial a_i} \sum_{j=1}^{p} u_j \sigma(a_j x + b_j) = x u_i \sigma'(a_i x + b_i)$

② asw $\nabla_a f_\theta(x) = (\sigma'(ax+b) \odot u)\, x = (\sigma'(ax+b) \odot (ux))$ ($\because x$ is scala)

$= \mathrm{diag}(\sigma'(ax+b))\, ux$ ($\because$ (ii) ~ ②)

※

P7



Data points
True Fn
Learned Fn after 2000 iterations
Learned Fn after 4000 iterations
Learned Fn after 6000 iterations
Learned Fn after 8000 iterations
Learned Fn after 10000 iterations