1. $L(y, x, \alpha) = (y - \hat{y})^2 = (y - \alpha x)^2$

$$R(\alpha) = E[(y - \alpha x)^2]$$
$$= E[(2x + \varepsilon - \alpha x)^2]$$
$$= E[((x(2-\alpha) + \varepsilon)^2]$$
$$= E[(x^2 \,2 - 4x^2\alpha + 4x^2 - 2x\alpha\varepsilon + 4x\varepsilon + \varepsilon^2)]$$
$$= E(x^2\alpha^2 - 4x^2\alpha + 4x^2)$$
$$= E(x^2(\alpha^2 - 4\alpha + 4))$$
$$= (\alpha^2 - 4\alpha + 4)\, E(x^2) \qquad | x \sim U(-2, 2)$$
$$= (\alpha^2 - 4\alpha + 4) \int_{-2}^{2} x^2 p(x)\, dx$$
$$= (\alpha^2 - 4\alpha + 4) \int_{-2}^{2} x^2 \frac{1}{2+2}$$
$$= \frac{(\alpha^2 - 4\alpha + 4)}{4} \Big/_{-2}^{2} \frac{x^3}{3}$$
$$= \frac{(\alpha^2 - 4\alpha + 4)}{4} \quad \frac{8}{3} + \frac{8}{3}$$
$$= \frac{4}{3}(\alpha^2 - 4\alpha + 4)$$

$$R(2.5) = \frac{4}{3}(2.5^2 - 4 \cdot 2.5 + 4)$$
$$= \frac{1}{3}$$

The numerical R(2.5 is about 2.5. With 100 samples you can get a very good estimate for the risk.

2.a) $P((x,y),\Theta) = \prod_{n=1}^{N} P(y_n=1)^{y_n} P(y_n=0)^{1-y_n}$

$$= \prod_{n=1}^{N} s(\Theta^T x)^{y_n} (1-\Theta^T x)^{1-y_n}$$

$$= \sum_{n=1}^{N} y_n \log(s(\Theta^T x)) + (1-y_n)\log(1-s(\Theta^T x))$$

b) $s'(z) = \frac{1}{1+e^{-z}} dz$

$$= (1+e^{-z})^{-1}$$

$$= (1+e^{-z})^{-2} e^{-z}$$

$$= \frac{e^{-z}}{(1+e^{-z})^2}$$

$$= \frac{e^{-z}}{(1+e^{-z})(1+e^{-z})}$$

$$= \frac{1}{(1+e^{-z})}\frac{e^{-z}}{(1+e^{-z})}$$

$$= \frac{1}{(1+e^{-z})}\left(1-\frac{1}{1+e^{-z}}\right)$$

$$= s(z)(1-s(z))$$

$$\frac{\partial L}{\partial \Theta} = -\sum_{n=1}^{N} \frac{y_n}{s(\Theta^T x_n)} \Delta s(\Theta^T x_n) + \frac{(1-y_n)}{1-s(\Theta^T x_n)} \Delta s(\Theta^T x_n)$$

$$= -\sum_{n=1}^{N}\left(\frac{y_n}{s(\Theta^T x_n)} + \frac{(1-y_n)}{1-s(\Theta^T x_n)}\right) \Delta s(\Theta^T x_n)$$

$$= -\sum_{n=1}^{N}\left(\frac{y_n}{s(\Theta^T x_n)} + \frac{(1-y_n)}{1-s(\Theta^T x_n)}\right) \cancel{s(\Theta^T x_n)}\,\cancel{(1-s(\Theta^T x_n)}\nabla(\Theta^T x_n)$$

$$= -\sum_{n=1}^{N}\left(\frac{y_n(1-s(\Theta^T x_n)+(1-y_n)s(\Theta^T x_n)}{\cancel{s(\Theta^T x_n)}\,\cancel{(1-s(\Theta^T x_n))}}\right) x_n$$

$$= -\sum_{n=1}^{N} (y_n - y_n s(\Theta^T x_n) + s(\Theta^T x_n) - y_n s(\Theta^T x_n)) x_n$$

$$= -\sum_{n=1}^{N} (y_n + s(\Theta^T x_n) - 2y_n s(\Theta^T x_n)) x_n$$

$$\approx -\sum_{n=1}^{N} (s(\Theta^T x_n) - y_n) x_n$$

c) $\log p(\theta) = \sum_{d=1}^{D} \log \mathcal{N}(0, \lambda^2)$

$= \sum_{d=1}^{D} \log \frac{1}{\sqrt{2\pi\lambda^2}} e^{-\frac{x^2}{2\lambda^2}}$

$= \sum_{d=1}^{D}$

?