1(d) By writing down the explicit gradient of $\log p(y|\hat{f}_x)$, make an argument that those training points x_i at which $|\hat{f}(x_i)| \gg 1$, those "far from the decision boundary" do almost not contribute to this estimate $\mathbb{E}_q[f(\cdot)]$.

Ans: $$\nabla_{f_x} \log(p(y|\hat{f}_x)) = \sum_{i=1}^{n} \nabla_{f_x} \log(\sigma(y_i f(x_i)))$$

$$= \sum_{i=1}^{n} \nabla_{f_x} \left[-\log(1+\exp(-y_i f(x_i)))\right]$$

$$\frac{\partial}{\partial f_i} - \log(1+\exp(-y_i f(x_i))) = \frac{+y_i \exp(-y_i f(x_i))}{1+\exp(-y_i f(x_i))}$$

$$y_i = -1 \Rightarrow \frac{(-1)\exp(f(x_i))}{1+\exp(f(x_i))} = \frac{-1}{1+\exp(-f(x_i))} = -\sigma(f(x_i))$$

$$y_i = 1 \Rightarrow \frac{(1)\exp(-f(x_i))}{1+\exp(-f(x_i))} = 1 - \frac{1}{1+\exp(-f(x_i))} = 1 - \sigma(f(x_i))$$

using mapping $$\left.\begin{array}{ll} y_i = -1 & ; \frac{y_i+1}{2} = 0 \\ y_i = 1 & ; \frac{y_i+1}{2} = 1 \end{array}\right\}$$

$$\frac{\partial}{\partial f_i} - \log(1+\exp(-y_i f(x_i))) = \frac{y_i+1}{2} - \sigma(f(x_i))$$

$$\nabla_i \log(p(y|\hat{f}_x)) = \frac{y_i+1}{2} - \sigma(f(x_i))$$

Now consider $|\hat{f}_x| >> 1$

$$\sigma(|\hat{f}_x|) = \frac{1}{1+\exp(-|\hat{f}_x|)} = 1 \quad \text{as} \quad \lim_{a\to\infty} \exp(-a) = 0$$

Thus when $y_i = 1$ and $|\hat{f}(x_i)| >> 1$ and $\hat{f}(x_i) > 0$

$$\nabla_i \log(p(y|\hat{f}_x)) = \frac{1+1}{2} - 1 = 0$$

Thus x_i will not contribute to $E_q[f(\cdot)]$

Similarly when $y_i = -1$ and $|\hat{f}(x_i)| >> 1$ ad $f(x_i) < 0$

$$\nabla_i \log(p(y|\hat{f}_x)) = \frac{-1+1}{2} - \frac{1}{1+\exp(-f(x_i))}$$

$$= 0 - \frac{1}{\underbrace{1+\exp(-f(x_i))}_{\lim \text{ as } -f(x_i)\to\infty}} = 0$$

thus those training point will not contribute to update equation. as the gradient of log likelihood

$$E_q[f(\cdot)] = m_\cdot + K_{\cdot x} \nabla \log p(y|\hat{f}_x)$$ those components is zero.