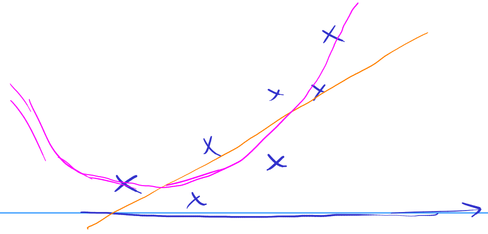
Bayesian model comparison
combination



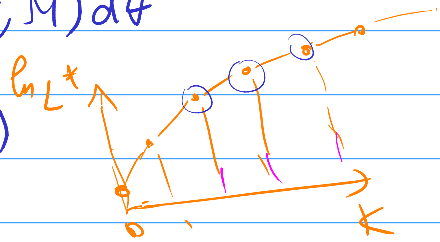
$$\left\{ \begin{matrix} M_1, & M_2, & M_3, & \dots \\ \theta_1 & \theta_2 & \theta_3 & \end{matrix} \right\} \quad D$$

$$p(M_i|D) \propto \cancel{p(M_i)}\, p(D|M_i)$$

$$p(\theta|D, M) = \frac{p(\theta|M_i)\; p(D|\theta, M_i)}{p(D|M_i)}$$

$$p(D|M) = \int p(\theta|M)\, p(D|\theta, M)\, d\theta$$

$$\theta^* = \theta_{MAP} \qquad \theta_{ML} \qquad \Big|\; L^* = p(D|\theta^*, M)$$

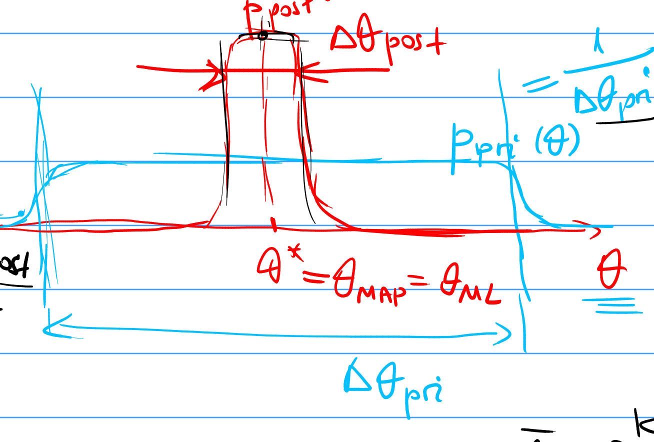


$$p_{post}(\theta) \propto p_{pri}(\theta)\cdot p(D|\theta) \propto p(D|\theta)$$

$$p(D|M) = \int \underbrace{p_{pri}(\theta)}_{}\, p(D|\theta)\, d\theta =$$

$$= \frac{1}{\Delta\theta_{pri}} \int\limits_{\Delta\theta_{pri}} p(D|\theta)\, d\theta =$$

$$p(D|\theta) \approx \begin{cases} p(D|\theta^*), & \text{в } \Delta\theta_{post} \\ 0, & \text{вне } \Delta\theta_{post} \end{cases}$$



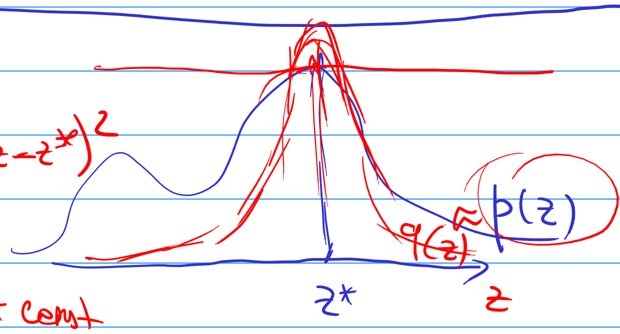
$$= \frac{1}{\Delta\theta_{pri}} \cdot p(D|\theta^*) \cdot \Delta\theta_{post}$$

$$\bar\theta \in \mathbb{R}^K$$

$$\ln p(D|M) = \ln p(D|\theta^*) + K \ln \frac{\Delta\theta_{post}}{\Delta\theta_{pri}} =$$

$$= \ln p(D|\theta^*) - K \ln \frac{\Delta\theta_{pri}}{\Delta\theta_{post}}$$

Лапласовская аппроксимация



$$\ln p(z) \approx p(z^*) + \cancel{\frac{\partial \ln p}{\partial z}\Big|_{z^*}}(z - z^*) + \frac{1}{2} \frac{\partial^2 \ln p}{\partial z^2}\Big|_{z^*} (z - z^*)^2$$

$$\ln q(z) = p(z^*) + \frac{1}{2} \frac{\partial^2 \ln p}{\partial z^2}\Big|_{z^*} (z - z^*)^2 + const$$

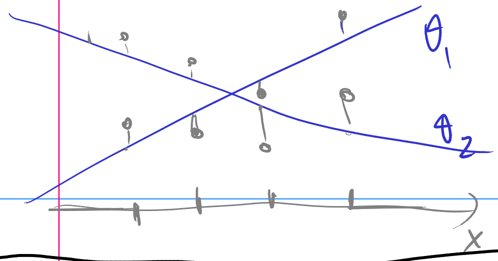
$$q(z) = \mathcal{N}\left(z \,\Big|\, z^*, -1\Big/\frac{\partial^2 \ln p}{\partial z^2}\Big|_{z^*}\right) \quad \leftarrow \text{Laplace approx.}$$

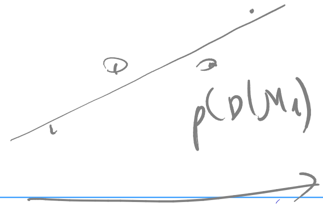
$$\ln p(\bar z) \approx \ln p(\bar z^*) + \nabla_{\bar z} \ln p\Big|_{z^*} (\bar z - \bar z^*) + \frac{1}{2} (\bar z - \bar z^*)^T \cdot H \cdot (\bar z - \bar z^*)$$

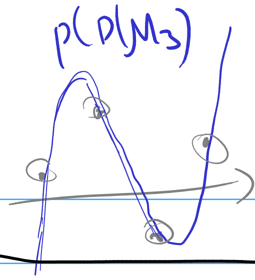
$$q(\bar z) = \mathcal{N}(\bar z \,|\, \bar z^*, -H^{-1})$$

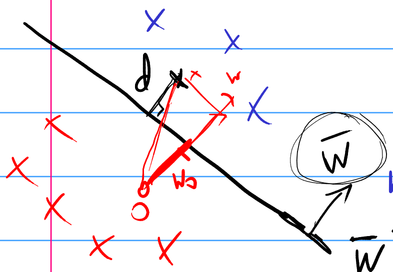
Hessian
гессиан

$$\left(\dots \frac{\partial^2 \ln p}{\partial z_i \partial z_j} \dots \right)$$





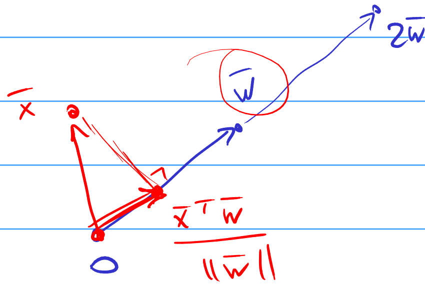


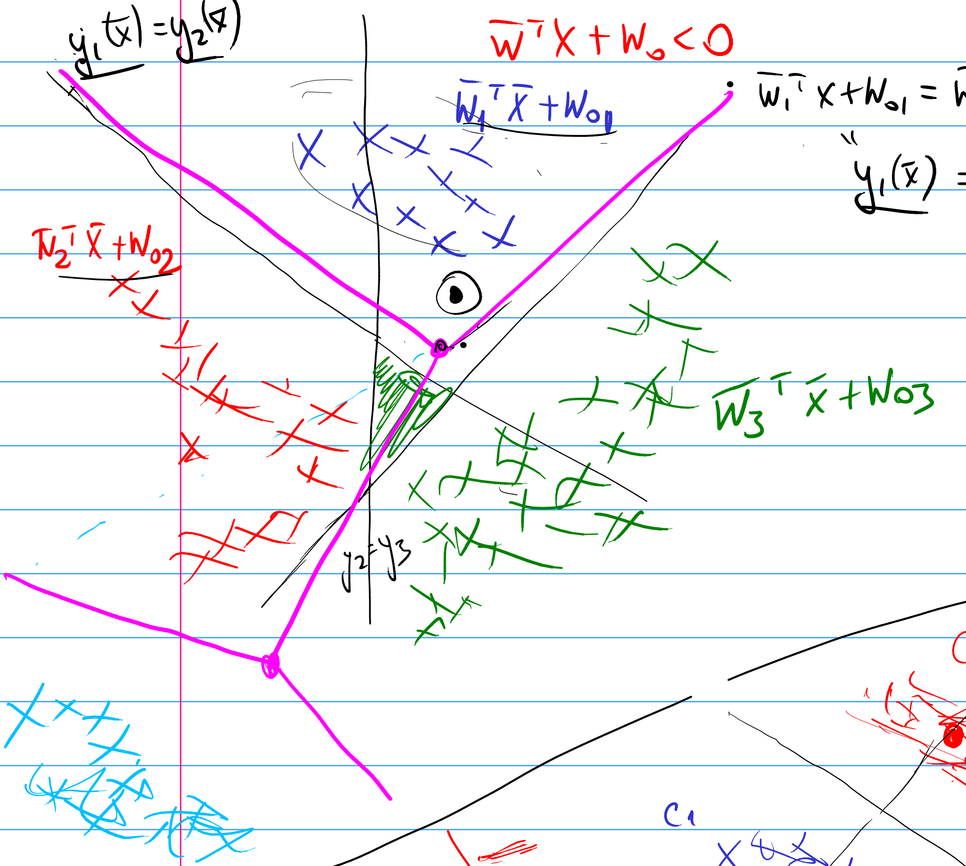


$$d(\bar{x}) = \frac{|\bar{w}^T\bar{x} + w_0|}{\|\bar{w}\|}$$

$\bar{w}^T\bar{x} + w_0 > 0$

$\bar{w}^T\bar{x} + w_0 = 0$





$\bar{w}_1^T x + w_{01} = \bar{w}_3^T\bar{x} + w_{03}$

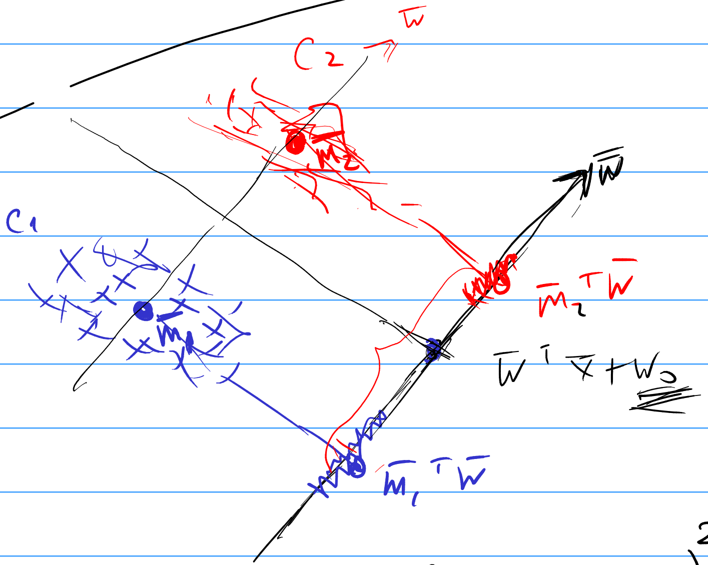
"

$y_1(\bar{x}) = y_3(\bar{x})$

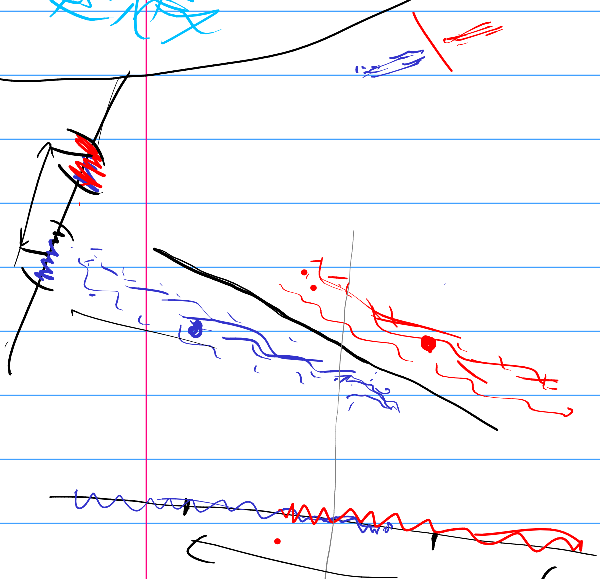
(min, +)

tropical math

pre-varieties



$$\left(\bar{w}^T(\bar{m}_1 - \bar{m}_2)\right)^2 \longrightarrow \max$$



$$\sum_{n \in C_1} (\bar{w}^T\bar{x}_n - \bar{w}^T\bar{m}_1)^2 + \sum_{n \in C_2} (\bar{w}^T\bar{x}_n - \bar{w}^T\bar{m}_2)^2 \longrightarrow \min$$

$$J(\bar{w}) = \frac{\left(\bar{w}^T(\bar{m}_1 - \bar{m}_2)\right)^2}{\sum\limits_{C_1} + \sum\limits_{C_2}} = \frac{\bar{w}^T (\bar{m}_1-\bar{m}_2)(\bar{m}_1-\bar{m}_2)^T \bar{w}}{\bar{w}^T\left(\sum\limits_{n \in C_1} (\bar{x}_n-\bar{m}_1)(\bar{x}_n-\bar{m}_1)^T + \sum\limits_{C_2} \ldots\right)\bar{w}} = \frac{\bar{w}^T S_B \bar{w}}{\bar{w}^T S_W \bar{w}} \xrightarrow[\bar{w}]{} \max$$

S_B - between-class cov.

S_W - within-class cov.

$$\sum_n (\bar{x}_n-\bar{m})(\bar{x}_n-\bar{m})^T = S = S_B + S_W$$

$$\nabla_{\bar{w}} J = \frac{2 S_B \bar{w} \cdot (\bar{w}^T S_W \bar{w}) - 2 S_W \bar{w} \cdot (\bar{w}^T S_B \bar{w})}{(\ldots)^2} = 0$$

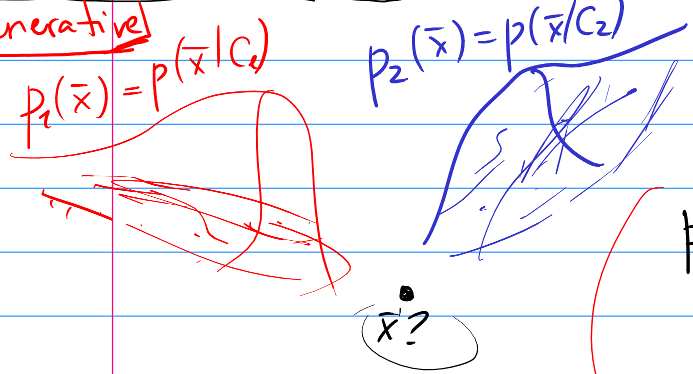
$A\bar{x} \propto B\bar{x}$

$$\cancel{(\bar{w}^T S_W \bar{w})} \cdot S_B \bar{w} \propto \cancel{(\bar{w}^T S_B \bar{w})} S_W \bar{w}$$

$$(\bar{m}_1-\bar{m}_2)(\bar{m}_1-\bar{m}_2)^T \bar{w} = S_B \bar{w} \propto S_W \bar{w} \propto \bar{m}_1 - \bar{m}_2$$

$$\bar{w} \propto S_W^{-1}(\bar{m}_1 - \bar{m}_2)$$

Generative



$p(\bar{x}/C_1)$ $\qquad$ $p(\bar{x}/C_2)$

$$p(C_1|\bar{x}) = \frac{p(C_1)p(\bar{x}|C_1)}{p(\bar{x})} = \qquad p(\bar{x}) = p(\bar{x},C_1) + p(\bar{x},C_2)$$

$= p(C_2|\bar{x})$

$$= \frac{p(C_1)\,p_1(\bar{x})}{p(C_1)p_1(\bar{x}) + p(C_2)p_2(\bar{x})} = 1/2$$

Optimal Bayesian classifier

$$p(\bar{x}|C_1) = \mathcal{N}(\bar{x}\,|\,\bar{\mu}_1, \Sigma_1)$$

$$p(C_1|\bar{x}) = p(C_2|\bar{x})$$

$$\ln p(C_1)p(\bar{x}/C_1) = \ln p(C_2)p(\bar{x}/C_2)$$

Σ^{-1}

$$\ln p(C_1) - \cancel{\frac{d}{2}\ln 2\pi} - \frac{1}{2}\ln\det\Sigma_1 - \frac{1}{2}(\bar{x}-\bar{\mu}_1)^T \Sigma_1^{-1}(\bar{x}-\bar{\mu}_1) =$$

Σ^{-1}

$$= \ln p(C_2) - \cancel{\frac{d}{2}\ln 2\pi} - \frac{1}{2}\ln\det\Sigma_2 - \frac{1}{2}(\bar{x}-\bar{\mu}_2)^T \Sigma_2^{-1}(\bar{x}-\bar{\mu}_2)$$

$$(\ldots\ldots) - \ln\frac{p(C_1)}{p(C_2)} = 0$$

$\Sigma_1 = \Sigma_2$ LDA

$\Sigma_1 \neq \Sigma_2$ QDA

Discriminative

$$p(C_1|\bar x) = \frac{p(C_1)p(\bar x|C_1)}{p(C_1)p(\bar x|C_1)+p(C_2)p(\bar x|C_2)} =$$

$$= \frac{1}{1+\frac{p(C_2)p(\bar x|C_2)}{p(C_1)p(\bar x|C_1)}} = \frac{1}{1+e^{-\ln\frac{p(C_1)p(\bar x|C_1)}{p(C_2)p(\bar x|C_2)}}}$$

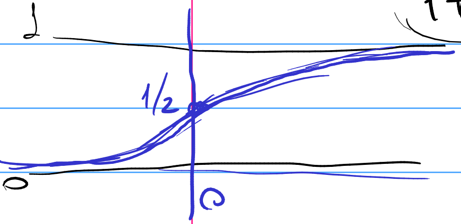
log odds $\in (-\infty, \infty)$

$\approx \bar w^T \bar x$

шансы odds $\frac{p(C_2|\bar x)}{p(C_1|\bar x)} \in [0, \infty]$ $\neq \bar w^T x$

$f(\bar w^T \bar x) \in [0,1]$ $\in(-\infty,\infty)$ $\Phi(a)$ $\int$

$$\sigma(a) = \frac{1}{1+e^{-a}} \qquad 1-\sigma(a) = 1-\frac{1}{1+e^{-a}} = \frac{e^{-a}}{1+e^{-a}} = \frac{1}{1+e^{a}} = \sigma(-a)$$



$$\sigma'(a) = \frac{+e^{-a}}{(1+e^{-a})^2} = \sigma(a)\cdot(1-\sigma(a))$$

$$p(D|\bar w) = \prod_{n=1}^N p(t_n|\bar w, \bar x_n) = \prod_{n=1}^N p(C_1|\bar w, \bar x_n)^{t_n}\, p(C_2|\bar w,\bar x_n)^{1-t_n} =$$

$\{(\bar x_n, t_n)\}_{n=1}^N$, $t_n \in \{0,1\}$

$$= \begin{cases} p(C_1|\bar w, \bar x_n), & t_n = 1 \\ p(C_2|\bar w, \bar x_n), & t_n = 0 \end{cases}$$

$$\nabla_{\bar w} \ln\sigma(\bar w^T \bar x_n) = \frac{\sigma_n(1-\sigma_n)}{\cancel{\sigma(\bar w^T \bar x_n)}}\bar x_n$$

$$= \prod_{n=1}^N \sigma(\bar w^T \bar x_n)^{t_n}\left(1-\sigma(\bar w^T \bar x_n)\right)^{1-t_n} \xrightarrow[\bar w]{} \max$$

$$\ln p(D|\bar w) = \sum_{n=1}^N \left(t_n \ln\sigma_n + (1-t_n)\ln(1-\sigma_n)\right)$$

$$\nabla_{\bar w} \ln p(D|\bar w) = \sum_{n=1}^N \left(t_n \cdot \frac{\cancel{\sigma_n}(1-\sigma_n)}{\cancel{\sigma_n}}\bar x_n - \frac{\sigma_n\cancel{(1-\sigma_n)}}{\cancel{1-\sigma_n}}\bar x_n(1-t_n)\right) =$$

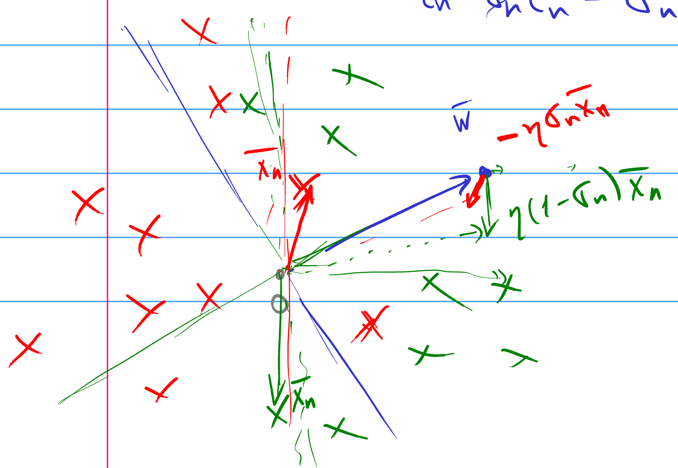
$\nabla_{\bar w}\nabla_{\bar w} \ln p(D|\bar w)$

$$= \sum_{n=1}^N \left(t_n(1-\sigma_n) - (1-t_n)\sigma_n\right)\bar x_n = \sum_{n=1}^N \left(t_n - \sigma_n(\bar w^T \bar x_n)\right)\bar x_n$$

$t_n - \cancel{\sigma_n t_n} - \sigma_n + \cancel{\sigma_n t_n}$

$\nabla_{\bar w} \ln p(d_n|\bar x)$

$-\alpha\bar w$



$\sigma_n > 1/2$

$t_n = 0$

$-\sigma_n \cdot \bar x_n$

$\sigma_n < 1/2$

$(1-\sigma_n)$

$$p(\bar{w} \mid D) \propto \underline{p(\bar{w})} \cdot p(D \mid \bar{w})$$

$$\sum_n \left(t_n \ln \sigma_n + (1-t_n)\ln(1-\sigma_n)\right) - \frac{\alpha}{2}\bar{w}^T\bar{w}$$

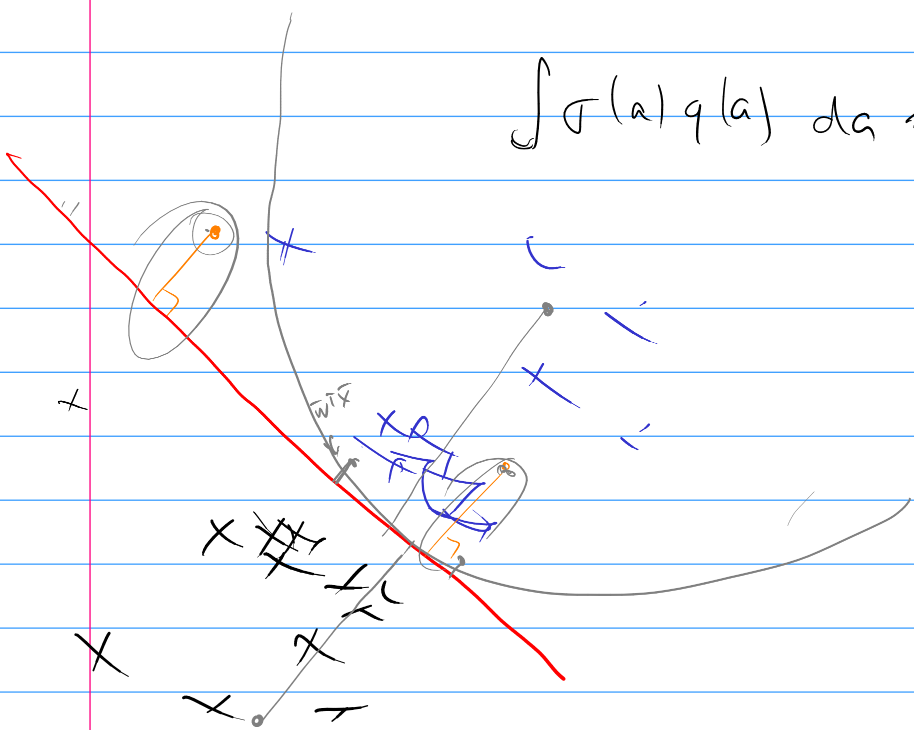
$$p(C_1 \mid \bar{x}, D) = \int p(C_1 \mid \bar{x}, \bar{w})\, p(\bar{w} \mid D)\, d\bar{w} =$$

$$= \int \sigma(\underline{\bar{w}^T\bar{x}}) \cdot \underbrace{p(\bar{w}) \cdot p(D \mid \bar{w})}_{\sim q(\bar{w}) \sim q(a)} \, d\bar{w}$$

$q(a)$ " $\bar{w}^T\bar{x}$ "

$q(\sim)$

$$\int \sigma(a)\, q(a)\, da \approx$$



$$\sigma(\kappa\, \bar{w}^T\bar{x})$$

$$\kappa = \frac{1}{\sqrt{1+\ldots^2}} < 1$$