① Logistic regression

$$p(C_1|\bar{x}) = \sigma(\bar{w}^T\bar{x}) = \frac{1}{1+e^{-\bar{w}^T\bar{x}}}$$

$$p(C_k|\bar{x}) = \frac{p(\bar{x}|C_k)p(C_k)}{\sum_i p(\bar{x}|C_i)p(C_i)} =$$

$$p(C_k|\bar{x}) = \frac{e^{\log p(\bar{x}|C_k)p(C_k)}}{\sum_{i=1}^{K} e^{\log p(\bar{x}|C_i)p(C_i)}} = \frac{e^{\log p(\bar{x}|C_k)p(C_k)+a}}{\sum_i e^{\log \text{—} + a}}$$

$$\bar{w}_k^T\bar{x} \sim \log p(\bar{x}|C_k)p(C_k)$$

$$\text{softmax}(a_1, a_2, \dots, a_K) = \left(\cdot \text{—} \frac{e^{a_k}}{\sum_i e^{a_i}} \text{— —} \right)$$

$$D = \{(\bar{x}_n, t_n)\}_{n=1}^{N}, \quad t_n = \begin{cases} 1, & \bar{x}_n \in C_1 \\ 0, & \bar{x}_n \in C_2 \end{cases} \qquad \bar{t}_n = (0, \dots, 0, 1, 0 \dots 0)$$

$\bar{x}_n \in C_k$; indices 1 … K, one-hot

$$p(D|\bar{w}) = \prod_{n=1}^{N} p(t_n|\bar{w}, \bar{x}_n) = \prod_{n=1}^{N} \begin{cases} \sigma(\bar{w}^T\bar{x}_n), & t_n = 1 \\ 1-\sigma(\bar{w}^T\bar{x}_n), & t_n = 0 \end{cases} =$$

$$= \prod_{n=1}^{N} \left[t_n \cdot \sigma(\bar{w}^T\bar{x}_n) + (1-t_n) \cdot (1-\sigma(\bar{w}^T\bar{x}_n)) \right]$$ (crossed out)

$$= \prod_{n=1}^{N} \sigma(\bar{w}^T\bar{x}_n)^{t_n} \cdot (1-\sigma(\bar{w}^T\bar{x}_n))^{1-t_n}$$

$$\sigma'(x) = \sigma(x)(1-\sigma(x))$$

② Max. likelihood in LR

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

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

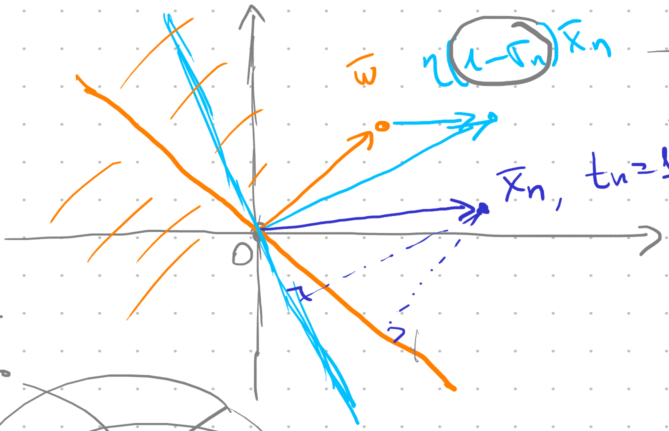
$\bar{W}^T\bar{X} + w_0$ $\quad \bar{w}^T\bar{x}$

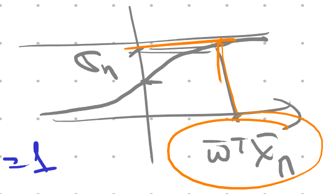
$$= X^T(\bar{t} - \bar{\sigma})$$

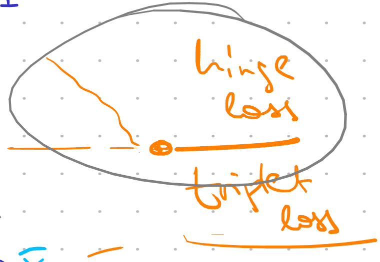
$t_n - t_n\sigma_n - \sigma_n + t_n\sigma_n$

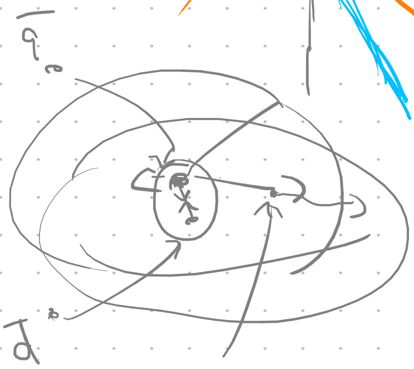
$$= \sum_n \left(t_n(1-\sigma_n) - (1-t_n)\sigma_n \right) \bar{x}_n = \sum_{n=1}^{N} \left(t_n - \sigma_n \right) \bar{x}_n =$$

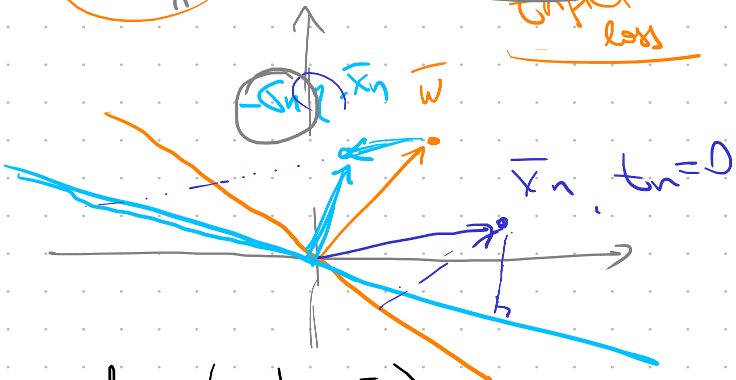
$t_n - \sigma(\bar{w}^T \bar{x}_n)$











SGD

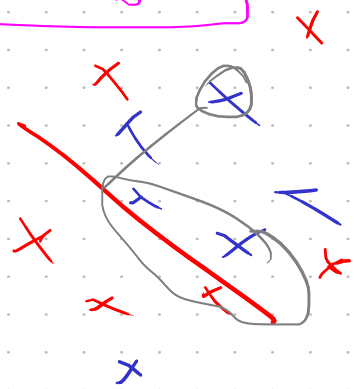
$$\bar{w} := \bar{w} + \eta \cdot \nabla_{\bar{w}} \log p(t_n | \bar{x}_n, \bar{w})$$

hessian

$$f(\bar{w}) \approx f(\bar{w}_0) + \nabla_{\bar{w}} f(\bar{w}_0)^T (\bar{w} - \bar{w}_0) + \frac{1}{2} (\bar{w} - \bar{w}_0)^T H_{\bar{w}_0} (\bar{w} - \bar{w}_0)$$

$$\nabla_{\bar{w}} f(\bar{w}_0) + H(\bar{w}_* - \bar{w}_0) = 0$$

$$\boxed{\bar{w}_* = \bar{w}_0 - H^{-1} \nabla_{\bar{w}} f(\bar{w}_0)}$$



$$\frac{\partial^2 \log p(D|\bar{w})}{\partial w_i \partial w_j} = \frac{\partial}{\partial w_i} \left(\sum_n \left(t_n - \sigma(\bar{w}^T \bar{x}_n) \right) x_{nj} \right) =$$

$$= \sum_n \left(-\sigma_n(1-\sigma_n) x_{ni} \right) x_{nj} = -\sum_n \sigma_n(1-\sigma_n) x_{ni} x_{nj}$$

ij:

$d \times d$

$$H = -\sum_n \sigma_n(1-\sigma_n) \bar{x}_n \bar{x}_n^T = -X^T \begin{pmatrix} \sigma_1(1-\sigma_1) & & 0 \\ & \ddots & \\ 0 & & \sigma_N(1-\sigma_N) \end{pmatrix} X$$

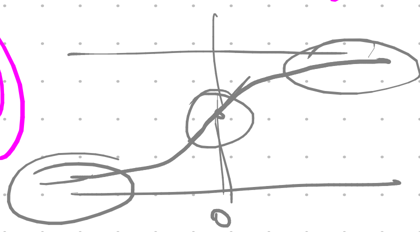
$d \times N$, $N \times N$, $\leftarrow R$, $N \times d$

$$H = -X^T R X \qquad X = \begin{pmatrix} \bar{x}_n \\ \vdots \end{pmatrix}$$

$$\bar{w}^{(k+1)} = \bar{w}^{(k)} + (X^T R X)^{-1} X^T \left(\bar{t} - \sigma(X \bar{w}^{(k)}) \right)$$

IRLS - iterative reweighted least squares

Lin. regs. $\bar{w}_* = (X^T X)^{-1} X^T \bar{y}$



$$p(\bar{w}|D) \propto p(\bar{w})\, p(D|\bar{w})$$

$$\log p(\bar{w}|D) = \text{Const} + \log p(D|\bar{w}) + \log p(\bar{w})$$

$$\log p(\bar{w}) = \text{const} - \frac{1}{2\sigma_0^2} \bar{w}^T \bar{w}$$

$$p(\bar{w}) = \mathcal{N}(\bar{w}|\bar{0}, \sigma_0^2 I)$$

$$\nabla \log p(\bar{w}|D) = \ldots - \frac{1}{\sigma_0^2}\bar{w}$$

$$H = \ldots - \frac{1}{\sigma_0^2} I$$

(3) Multiclass log.-regr.

$$\bar{t}_n = (0 \ldots 1 \ldots 0)$$

$$y_{nk} = \frac{e^{\bar{w}_k^T \bar{x}_n}}{\sum_j e^{\bar{w}_j^T \bar{x}_n}}, \quad \bar{w}_k^T \bar{x}_n = a_k, \quad \bar{w}_j^T \bar{x}_n = a_j$$

$$p(D|\bar{w}) = \prod_{n=1}^{N} \prod_{k=1}^{K} p(C_k|\bar{x}_n, \bar{w})^{t_{nk}}$$

$$\log p(D|\bar{w}) = \sum_n \sum_k t_{nk} \log y_{nk}$$

$$\frac{\partial y_{nk}}{\partial a_j} =$$

(1) $k \neq j$

$$= -\frac{e^{a_k} e^{a_j}}{(\sum_s e^{a_s})^2} = -y_{nk} y_{nj}$$

(2) $k = j$

$$\frac{e^{a_k}(\sum_s e^{a_s}) - e^{a_j} e^{a_k}}{(\sum_s e^{a_s})^2} = y_{nk}(1 - y_{nj})$$

$$\nabla_{\bar{w}_j} \log p(D|\bar{w}) = \sum_n \sum_k t_{nk} \frac{1}{y_{nk}} \cdot y_{nk}([k=j] - y_{nj})\bar{x}_n$$

$$= \sum_n \sum_k t_{nk}([k=j] - y_{nj})\bar{x}_n$$

$$= \sum_n \Big[t_{nj}(1 - y_{nj}) - \sum_{k \neq j} t_{nk} y_{nj} \Big] \bar{x}_n$$

$$= \sum_n \Big(t_{nj} - \big(\sum_k t_{nk}\big) y_{nj} \Big) \bar{x}_n = \quad \left(\sum_k t_{nk} = 1\right)$$

$$\frac{\partial}{\partial w_{is}}\left(\frac{\partial}{\partial w_{jl}} \ldots\right) = \sum_n (t_{nj} - y_{nj})\bar{x}_n$$

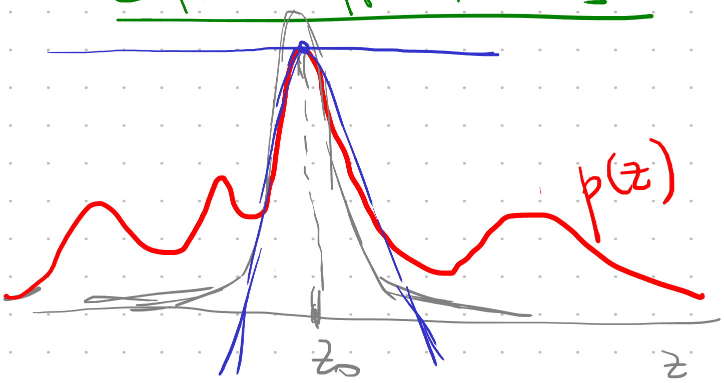
$$\frac{\partial}{\partial \bar{w}_i}\left(\nabla_{\bar{w}_j} \ldots\right) = \frac{\partial}{\partial \bar{w}_i}\left(\sum_n (t_{nj} - y_{nj})\bar{x}_n\right) = \begin{cases} -\sum_n y_{ni}(1 - y_{ni})\bar{x}_n \\ +\sum_n y_{nj} y_{ni} \bar{x}_n \end{cases}$$

4 Predictive distribution in log-regs.

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

$$= \int \sigma(\bar{w}^T\bar{x}) \cdot p(\bar{w}|D)\, d\bar{w} = E_{p(\bar{w}|D)}\left[\sigma(\bar{w}^T\bar{x})\right]$$

Laplace approximation



$$\log p(z) \approx \log p(z_0) + \left.\frac{\partial \log p}{\partial z}\right|_{z_0}(z-z_0)$$

$$+ \frac{1}{2}\left.\left(\frac{\partial^2 \log p}{\partial z^2}\right)\right|_{z_0}(z-z_0)^2$$

$$= \log p(z_0) - \frac{1}{2} A \cdot (z-z_0)^2$$

$$A = -\left(\frac{\partial^2}{\partial \ldots}\right)$$

$$p(z) \approx p(z_0) \cdot e^{-\frac{1}{2}A(z-z_0)^2} \approx \mathcal{N}\left(z \,|\, z_0, \frac{1}{A}\right)$$

$$\log p(\bar{z}) \approx \log p(\bar{z}_0) - \frac{1}{2}(\bar{z}-\bar{z}_0)^T A (\bar{z}-\bar{z}_0)$$

$$A = -\nabla\nabla \log p(\bar{z})$$

$$p(C_1|D, \bar{x}) = \int \sigma(\bar{w}^T\bar{x})\, p(\bar{w}|D)\, d\bar{w} \approx [\bar{w}_{MAP}] \approx$$

$$\approx \int \sigma(\underbrace{\bar{w}^T\bar{x}}_{\text{"}a\text{"}}) \cdot \mathcal{N}(\bar{w} \,|\, \bar{w}_{MAP}, \Sigma_N)\, d\bar{w}$$

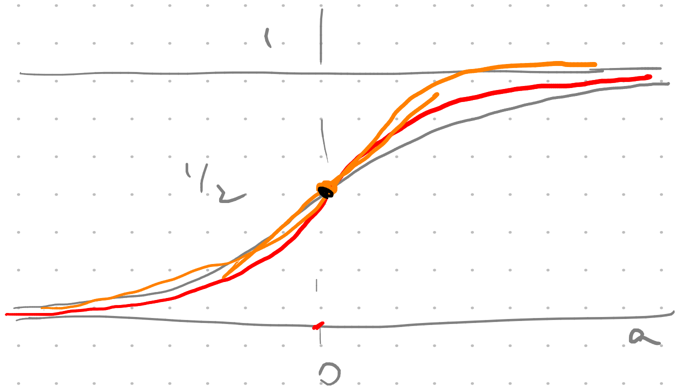
$$\sigma(\bar{w}^T\bar{x}) = \int_{-\infty}^{\infty} \sigma(a) \cdot \delta(\bar{w}^T\bar{x} - a)\, da$$

$$= \int_{-\infty}^{\infty} \sigma(a) \left(\int \delta(\bar{w}^T\bar{x} - a)\, \mathcal{N}(\bar{w} \,|\, \bar{w}_{MAP}, \Sigma_N)\, d\bar{w} \right) da$$

$$= \int_{-\infty}^{\infty} \sigma(a)\, \mathcal{N}(a \mid \mu_a, \sigma_a^2)\, da \approx$$

$\sigma_a^2 = \bar{x}^T \Sigma_N \bar{x}$

$\mu_a = \bar{w}_{MAP}^T \bar{x}$



probit function

$$\Phi(b) = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx$$

$$\Phi(\lambda a) \approx \sigma(a)$$

$$\Phi\left(\sqrt{\tfrac{\pi}{8}}\, a\right) \approx \sigma(a)$$

$$\approx \int_{-\infty}^{\infty} \int_{-\infty}^{\lambda a} \mathcal{N}(x \mid 0, 1) \cdot \mathcal{N}(a \mid \mu_a, \sigma_a^2)\, dx\, da = [\text{упражнение}]$$

$$= \Phi\left(\frac{\mu_a}{\sqrt{\sigma_a^2 + \frac{1}{\lambda^2}}}\right) = \Phi\left(\frac{\lambda \mu_a}{\sqrt{1 + \lambda^2 \sigma_a^2}}\right) \approx$$

$$\approx \sigma\left(\frac{\mu_a = \bar{w}_{MAP}^T \bar{x}}{\sqrt{1 + \frac{\pi}{8} \cdot \sigma_a^2}}\right)$$