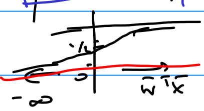
$\sigma(-a) = 1 - \sigma(a)$

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

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

$a \approx \bar{w}^T\bar{x}$

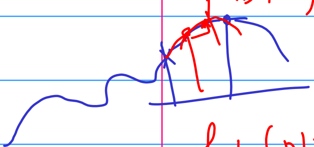


$D = \{(\bar{x}_n, t_n)\}^N \qquad t_n = \begin{cases} 0, & x \in C_2 \\ 1, & x \in C_1 \end{cases}$

$$p(t_n|\bar{x}_n, \bar{w}) = \begin{cases} \frac{1}{1+e^{-\bar{w}^T\bar{x}_n}}, & t_n = 1 \\ 1 - \frac{1}{1+e^{-\bar{w}^T\bar{x}}}, & t_n = 0 \end{cases} = \begin{cases} \sigma(\bar{w}^T\bar{x}_n), & t_n = 1 \\ 1-\sigma(\bar{w}^T\bar{x}_n), & t_n = 0 \end{cases}$$

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

$$p(D|\bar{w}) = \prod_D p(t_n|\bar{x}_n, \bar{w}) = \prod_D \sigma(\bar{w}^T\bar{x}_n)^{t_n}(1-\sigma(\bar{w}^T\bar{x}_n))^{1-t_n}$$



$$\ln p(D|\bar{w}) = \sum_D \left[t_n \ln\sigma(\bar{w}^T\bar{x}_n) + (1-t_n)\ln(1-\sigma(\bar{w}^T\bar{x}_n))\right] \to \max$$

$t_n \cdot \sigma(.) + (1-t_n)\cdot(1-\sigma)$

$$\nabla_{\bar{w}} \ln p(D|\bar{w}) = \sum_n \left[t_n \cdot (1-\sigma(\bar{w}^T\bar{x}_n))\bar{x}_n - (1-t_n)\sigma(\bar{w}^T\bar{x}_n)\bar{x}_n\right] = \sum_n (t_n - \sigma(\bar{w}^T\bar{x}_n))\bar{x}_n$$

perceptron

$\cancel{t_n}(1-\cancel{\sigma}) - (1-\cancel{t_n})\sigma = t_n - \sigma$

$y_n(\bar{x}_n) = p(C_1|\bar{x}_n, \bar{w})$

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

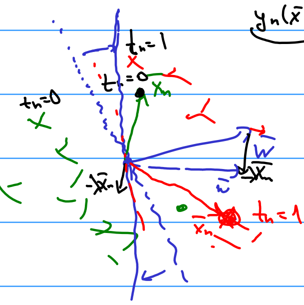
$(X^T \Sigma X)^{-1}(X^T \bar{t})$ IRLS

$\begin{pmatrix} \sigma(1-\sigma) & \\ & \ddots \\ & & \sigma(1-\sigma) \end{pmatrix}$

$\bar{w} := \bar{w} + \eta \cdot \nabla_{\bar{w}} \ln p$

$+ \lambda \bar{w}^T\bar{w}$

$+ \lambda \sum |w_i|$



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

(1) $\ln p(D|\bar{w}) \to \max$

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

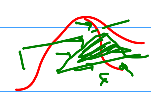
$p(\bar{w}) = N(\bar{w}|\mu_0, \Sigma_0)$

$\ln p(\bar{w}) = \lambda \sum |w_i| + \text{const}$

(3) $p(C_1|\bar{x}, D) = \int p(C_1|\bar{x}, \bar{w}) \cdot p(\bar{w}|D)\, d\bar{w} \propto \int \sigma(\bar{w}^T\bar{x}) \cdot \boxed{N(\bar{w}|\mu_0, \Sigma_0)\left[\prod_n \sigma_n^{t_n}(1-\sigma_n)^{1-t_n}\right]} d\bar{w}$

$N(\bar{w}|\mu_N, \Sigma_N)$

$\approx \int \sigma(\underbrace{\bar{w}^T\bar{x}}_{= a}) \cdot N(\bar{w}|\mu_N, \Sigma_N)\, d\bar{w}$

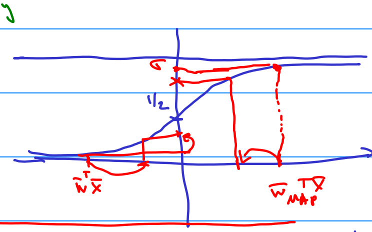


$= \int \sigma(a) N(a|-)\, da \approx \iint^a N(y|.) N(a|..)\, dy\, da = \int N(y|..)\, dy = \Phi(..) \approx \sigma(..)$

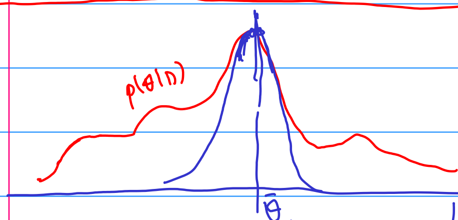
$\approx \Phi(a) = \int_{-\infty}^a N(y)\, dy$

$f(x) \approx c_0 + c_1(x-x_0) + c_2(x-x_0)^2$

$$p(C_1|\bar{x}, D) = \sigma\left(\frac{1}{\sqrt{1+\frac{\pi}{8}\sigma^2}}(\bar{w}_{MAP}^T\bar{x})\right)$$



$\boxed{\ln p(D) \approx \ln p(D|\theta_{MAP}) - \frac{1}{2}M\ln N \quad \text{BIC} \quad \text{Schwarz}}$

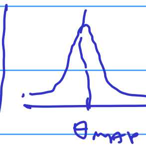


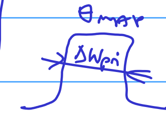
$p(\theta|D) \propto p(\theta)p(D|\theta) \approx N = \dots e^{-\frac{1}{2}(\bar{\theta}-\bar{\mu})^T\Sigma^{-1}(\bar{\theta}-\bar{\mu})}$

$\ln p(\theta|D) \approx C - \frac{1}{2}(\bar{\theta}-\bar{\theta}_{MAP})^T A(\bar{\theta}-\bar{\theta}_{MAP})$

Laplace approximation

$\boxed{p(\theta|D) \approx c\, e^{-\frac{1}{2}(\bar{\theta}-\bar{\theta}_{MAP})^T A(\bar{\theta}-\bar{\theta}_{MAP})}}$

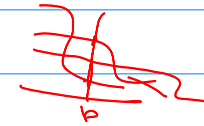




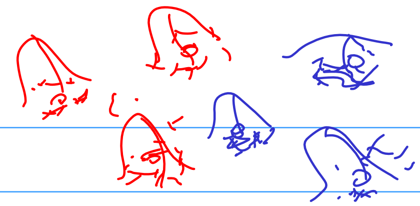
$M_1, \dots, M_K \qquad p(M_i|D) \propto p(D|M_i)p(M_i)$

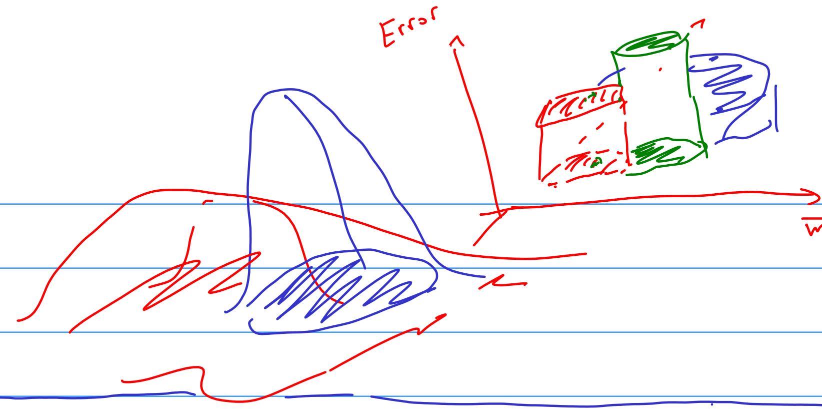
$p(D|M_i) = \int p(\theta|M_i)p(D|\theta, M_i)\, d\theta$

$\ln p(D) \approx \ln p(D|\theta_{MAP}) - \ln\frac{\Delta W_{pri}}{\Delta W_{post}}$

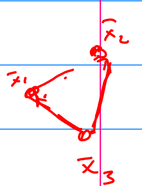


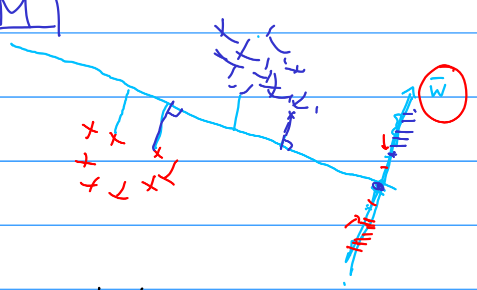
$$\ln p(D) \approx \ln p(D|\theta_{MAP}) + \underbrace{\ln p(\theta_{MAP}) + \frac{M}{2}\ln 2\pi - \frac{1}{2}\ln\det A}_{\text{Ockham's factor}}$$

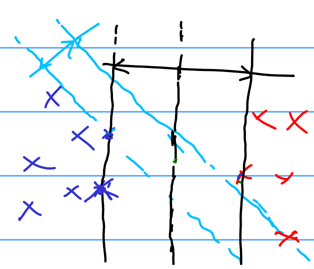


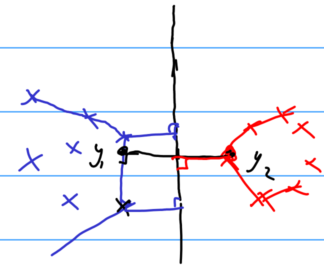


SVM









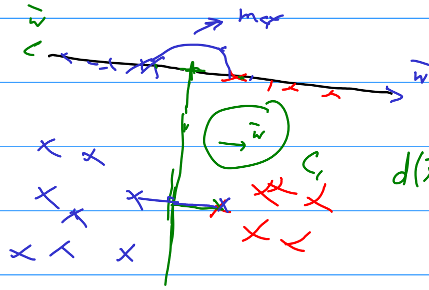
$C_1 = \{\bar x_n\}_{n: t_n = 1} \qquad C_2 = \{\bar x_n\}_{n: t_n = -1}$

$\| \underset{\in Conv(C_1)}{\bar y_1} - \underset{\in Conv(C_2)}{\bar y_2} \|^2 \longrightarrow \min$

$\left\| \sum_{t_n=1} \alpha_n \bar x_n - \sum_{t_n=-1} \alpha_n \bar x_n \right\|^2 \xrightarrow{\bar\alpha} \min$

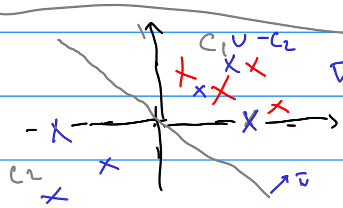
quadratic programming $\quad \begin{cases} \alpha_n \ge 0 \quad \forall n \\ \sum_{t_n=1} \alpha_n = \sum_{t_n=-1} \alpha_n = 1 \end{cases}$

SVM



$d(\bar x_n, \text{"}w\text{"}) = \frac{|\bar w^T \bar x_n + w_0|}{\|\bar w\|}$

$\max_{\bar w} \min_{\substack{n: t_n = 1 \\ m: t_m = -1}} |\bar w^T \bar x_n - \bar w^T \bar x_m|$



$\bar w^T \bar x > 0 \quad x \in C_1$

$\bar w^T \bar x < 0 \quad x \in C_2$

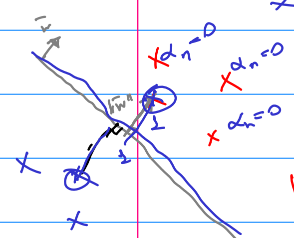
$\Downarrow$

$\bar w^T (-\bar x) > 0$

$\bar x \in C_2$

$\max_w \min_n \frac{|\bar w^T \bar x_n + w_0|}{\|\bar w\|} = \max_{\bar w} \min_n \frac{t_n (\bar w^T \bar x_n + w_0)}{\|\bar w\|}$

$\forall n \begin{cases} \bar w^T \bar x_n + w_0 > 0, & t_n = 1 \\ \bar w^T \bar x_n + w_0 < 0, & t_n = -1 \end{cases} \Leftrightarrow \forall n \quad t_n (\bar w^T \bar x_n + w_0) > 0$



$\max_{\bar w} \left(\frac{1}{\|\bar w\|} \cdot \underbrace{\min_n t_n (\bar w^T \bar x_n + w_0)}_{=1} \right)$

SVM $\quad \max_{\bar w} \frac{1}{\|\bar w\|} = \min_{\bar w} \|\bar w\| = \min_{\bar w} \|\bar w\|^2$

$\forall n \quad t_n (\bar w^T \bar x_n + w_0) \ge 1$

$L(\bar w, w_0, \bar\alpha) = \frac{1}{2} \underbrace{\|\bar w\|^2}_{\bar w^T \bar w} - \sum_n \alpha_n [t_n (\bar w^T \bar x_n + w_0) - 1] = \sum_n \alpha_n t_n \bar w^T \bar x_n - w_0 \cancel{\sum_n \alpha_n t_n} + \sum_n \alpha_n, \quad \alpha_n \ge 0$

$\frac{\partial L}{\partial w_k} = w_k - \sum_n \alpha_n t_n x_k = 0 \qquad w_k = \sum_n \alpha_n t_n x_k$

$\boxed{\bar w = \sum_n \alpha_n t_n \bar x_n}$

$\frac{\partial L}{\partial w_0} = -\sum_n \alpha_n t_n = 0 \qquad \sum_n \alpha_n t_n = 0$

$L = \frac{1}{2} \left(\sum \alpha_n t_n \bar x_n\right)^T \left(\sum \alpha_n t_n \bar x_n\right) - \sum_n \alpha_n t_n \bar x_n^T \left(\sum_m \alpha_m t_m \bar x_m\right) + \sum_n \alpha_n$

$= \sum_{n,m} \alpha_n \alpha_m t_n t_m \bar x_n^T \bar x_m \qquad \sum_{n,m} \alpha_n \alpha_m \ldots$

SVM (N)

$L(\bar\alpha) = \sum_n \alpha_n - \frac{1}{2} \sum_n \sum_m \alpha_n \alpha_m t_n t_m \bar x_n^T \bar x_m \xrightarrow{\bar\alpha} \min$

$\alpha_n \ge 0 \qquad \sum_n \alpha_n t_n = 0$

$\bar w = \sum_n \alpha_n t_n \bar x_n, \quad w_0$

$y(\bar x) = \sum_n \alpha_n t_n \bar x^T \bar x_n + w_0$