Updates:

e) X, Y are independent:

		Y		
		1	2	3
X	1	0.1	0.15	0.25
	2	0.05	0.3	0.15

$P(XY) = P(X)P(Y)$

$\Rightarrow P(X=x_i, Y=y_i) = P(X=x_i)\,P(Y=y_i)$

$\Rightarrow$ let $x_i = 1, y_i = 1$, $P(X=1, Y=1) = 0.1$ ~~P(X)~~ = $P(X=1) = 0.5$

$P(Y=1) = 0.15$

$\Rightarrow P(X=1)P(Y=1) = 0.5 \times 0.15 = 0.075 \neq P(X=1, Y=1)$.

X Y are independent.

☆ What's difference between Covarience and Correlations?

In short, correlation is a normalized version of covariance.

☆ What's relationship between Correlations and independence?

Independence $\Rightarrow$ uncorrelatedness

correlations only ~~capture~~ linear dependence.

correlations is synonymous with dependence

☆ what's meaning of correlations and covariance.

corelationce: distance of dependences.

Ex 8-1

a) $\vec{y} = \vec{X}\vec{w} + \vec{\varepsilon}$ $\qquad p(\vec{\varepsilon}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\vec{\varepsilon}^{\,2}\right\}$

$$p(\vec{y}\,|\,\vec{X},\vec{w}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(\vec{y}-\vec{X}\vec{w})^2\right\}$$

$d_i = (x_{i,1}, \cdots, x_{i,M}, y_i)^T \in \vec{D}$.

$$L(\vec{w}) = p(\vec{D}|\vec{w}) = p(\vec{X},\vec{y}\,|\,\vec{w}) = p(\vec{y}|\vec{X},\vec{w}) \cdot p(\vec{X}|\vec{w})$$

$$= p(\vec{y}|\vec{x},\vec{w}) \cdot p(\vec{x})$$

($\vec{x}$ independent with $\vec{w}$).

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(\vec{y}-\vec{x}\vec{w})^2\right\} p(\vec{x})$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(\vec{y}^T\vec{y} - 2\vec{w}^T\vec{X}^T\vec{y} + \vec{w}^T\vec{X}^T\vec{x}\,\vec{w})\right\} \cdot p(x).$$

$$\frac{\partial \ln L(\vec{w})}{\partial \vec{w}} = \frac{\partial\left(-\frac{1}{2\sigma^2}(\vec{y}^T\vec{y} - 2\vec{w}^T\vec{x}^T\vec{y} + \vec{w}^T\vec{x}^T\vec{x}\,\vec{w})\right)}{\partial \vec{w}}$$

$$= \left(-2\vec{X}^T\vec{y} + 2\vec{X}^T\vec{x}\,\vec{w}\right)\frac{1}{2\sigma^2} = \frac{1}{\sigma^2}(\vec{x}^T\vec{y} - \vec{x}^T\vec{x}\vec{w})$$

BRUNNEN

b) $L(\vec{w}) = P(\vec{w})\, P(\vec{D}|\vec{w})$ (Maximum - posteriori - estimator)

$$\ln L(\vec{w}) = \ln P(\vec{w}) + \ln P(\vec{D}|\vec{w}).$$

$$\frac{\partial \ln L(\vec{w})}{\partial \vec{w}} = \frac{\partial \ln P(\vec{w})}{\partial \vec{w}} + \frac{1}{\sigma^2}\left(\vec{X}^T\vec{y} - \vec{X}^T\vec{X}\vec{w}\right).$$

$$= \frac{\partial}{\partial \vec{w}}\left(\ln\left(\frac{1}{\sqrt{2\pi\gamma^2}^M} \exp\left\{-\frac{1}{2\alpha^2}\vec{w}^T\vec{w}\right\}\right)\right) + \cdots$$

$$= 0 + \frac{\partial}{\partial \vec{w}}\left(-\frac{1}{2\alpha^2}\vec{w}^T\vec{w}\right) + \cdots$$

$$= -\frac{1}{\alpha^2}\vec{w} + \frac{1}{\sigma^2}\left(\vec{X}^T\vec{y} - \vec{X}^T\vec{X}\vec{w}\right) = 0$$

$$\frac{\sigma^2}{\alpha^2}\vec{w} = \left(\vec{X}^T\vec{y} - \vec{X}^T\vec{X}\vec{w}\right).$$

$$\Rightarrow \left(\vec{X}^T\vec{X} + \frac{\sigma^2}{\alpha^2} I\right)\vec{w} = \vec{X}^T\vec{y}$$

$$\Rightarrow \hat{\vec{w}} = \left(\vec{X}^T\vec{X} + \frac{\sigma^2}{\alpha^2} I\right)^{-1}\vec{X}^T\vec{y}.$$

$$\hat{\vec{w}}_{pen} = \left(\vec{X}^T\vec{X} + \lambda I\right)^{-1}\vec{X}^T\vec{y}.$$

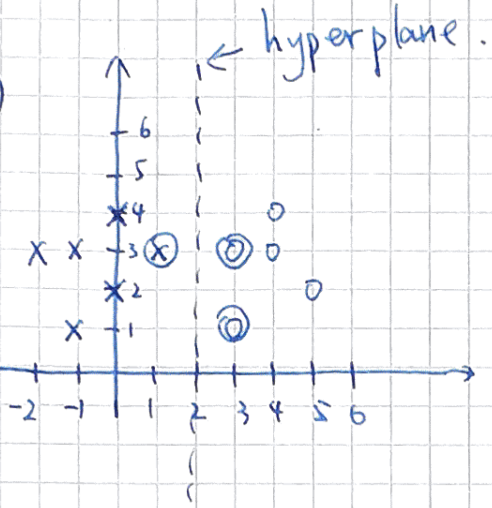
I.e. $\lambda = \frac{\sigma^2}{\alpha^2}$ MAP correspods to regularized cost.

$$\text{cost}(\vec{w}) = \left(\vec{Y} - f(\vec{X}, \vec{w})\right)^2 + \lambda\vec{w}^T\vec{w}$$

BRUNNEN

Ex 8-2

a)



⊗ ⊙ ⊙ are supporting vectors.

b) hyperplane $x = 2$.

$h_1 = -2 + x + 0 \cdot y$.

$h_2 = 2 - x + 0 \cdot y$.

$w_1 = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}$ $\quad w_2 = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$

for $\otimes = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ has 1.

$h_1 = -2 + 1 + 0 = -1$ incorrect.

$h_2 = 2 - 1 + 0 = 1$ correct.

Thus. $w = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$.

---

c) margin $C = \frac{1}{\|\tilde{\vec{w}}\|}$ where $\tilde{\vec{w}}$ is optimal $\vec{w}$ without w_0 bias.

Thus $\tilde{\vec{w}} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$. $\quad C = \frac{1}{\sqrt{(-1)^2 + 0^2}} = 1$. is the margin.
