Linear regression

$[\underline{x} - \underline{q}][\underline{x} -$

Consider $\boxed{t = \underline{\phi}^T(\underline{x})\,\underline{w} + \epsilon}$

"t" is a linear function of the vector $\underline{w}$

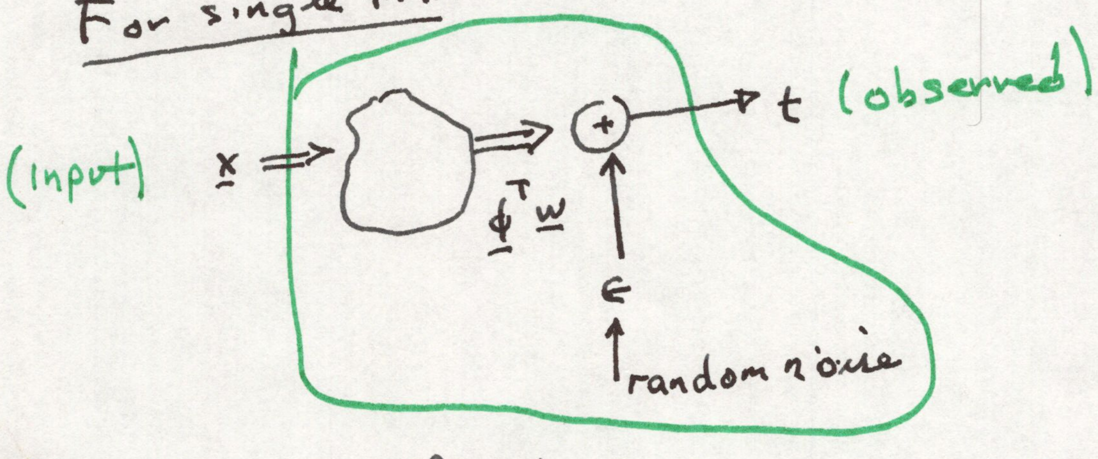
where $\underline{X} = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_Q \end{Bmatrix}$ $\quad \underline{\phi}(\underline{x}) = \begin{Bmatrix} \phi_1(\underline{x}) \\ \phi_2(\underline{x}) \\ \vdots \\ \phi_M(\underline{x}) \end{Bmatrix}$ $\Leftarrow$ given

the unknown

$$\underline{W} = \begin{Bmatrix} w_1 \\ w_2 \\ \\ w_{M-1} \\ w_M \end{Bmatrix}$$

N # trial
M # order

For single trial



$\epsilon : f_\epsilon(\epsilon)$

Problem

Given ① $\underline{\phi}^T \underline{w}$ ③ $t = \underline{\phi}^T \underline{w} + \epsilon$

② $\epsilon : N(0, \sigma^2)$

find $f_T(t)$

$$\int_{-\infty}^{t} f_T(q)\,dq = \int_{-\infty}^{(\epsilon)} f_\epsilon(\beta)\,d\beta$$

however $\epsilon = t - \underline{\phi}^T \underline{w}$

therefore

$$\int_{-\infty}^{t} f_T(q)\,dq = \int_{-\infty}^{t-\underline{\phi}^T\underline{w}} f_\epsilon(\beta)\,d\beta$$

$$\frac{\partial}{\partial t}\left[\int_{-\infty}^{t} f_T(q)\,dq\right] = \frac{\partial}{\partial t}\left[\int_{-\infty}^{t-\underline{\phi}^T\underline{w}} f_\epsilon(\beta)\,d\beta\right]$$

$$\boxed{f_T(t) = f_\epsilon(t - \underline{\phi}^T \underline{w})}$$

Recall

$\epsilon : N(0, \sigma^2)$

hence

$$f_\epsilon(\epsilon) = \frac{e^{-\left[\frac{\epsilon^2}{2\sigma^2}\right]}}{\sqrt{2\pi\sigma^2}}$$

The pdf of t is

$$\boxed{f_T(t) = \frac{e^{-\frac{(t - \underline{\phi}^T \underline{w})^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}}$$

$$\boxed{\hat{N}(t \mid \underline{\phi}^T \underline{w}) \equiv f_T(t)}$$

$E(t) = \underline{\phi}^T \underline{w}$

$E(t^2) - E(t)^2 = \sigma^2.$

N-trials (~~indepents~~ independent)

$$\underline{t} = \{t_1, t_2 \ldots, t_N\}$$

$$\underline{X} = \{\underline{x}_1, \underline{x}_2, \cdots \underline{x}_N\}$$

$$p(\underline{t} \mid \underline{X}, \underline{w}) = \prod_{n=1}^{N} \hat{N}\left(t_n \mid \underline{\phi}^T(\underline{x}_n)\underline{w}\right)$$

$$p(t \mid \underline{X}, \underline{w}) = \frac{e^{-\sum_{n=1}^{N} \frac{(t_n - \underline{\phi}^T(x_n)\underline{w})^2}{2\sigma^2}}}{\left(\sqrt{2\pi\sigma^2}\right)^N}$$

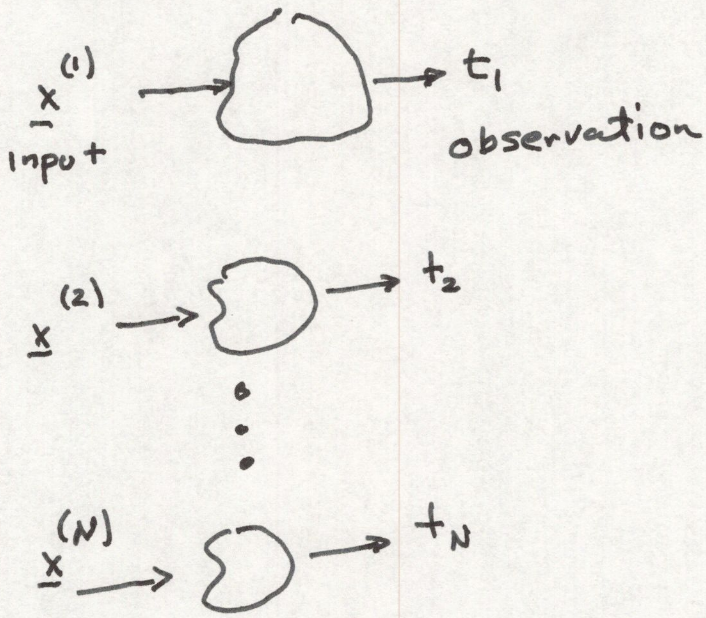
Simplified

$$[\Phi] = \begin{bmatrix} \phi_1(\underline{x}_1) & \cdots & \phi_M(\underline{x}_1) \\ \vdots & & \vdots \\ \phi_1(\underline{x}_N) & \cdots & \phi_M(\underline{x}_N) \end{bmatrix}$$

$$[\Phi] = \begin{bmatrix} \underline{\phi}^T(x_1) \\ \underline{\phi}^T(x_2) \\ \vdots \\ \underline{\phi}^T(x_N) \end{bmatrix}$$

1

Example



trial 1

trial 2

trail N

Single trial:

Given $\underline{x}^{(j)} = \begin{bmatrix} 1 \\ x_2^{(j)} \\ x_3^{(j)} \end{bmatrix}$ & t_j

Model

$$t_j = \underline{\phi}^T(\underline{x}^{(j)})\,\underline{w} + \epsilon^{(j)}$$

define/choose Basis

$$\underline{\phi}(\underline{x}^{(j)}) = \begin{bmatrix} \phi_1(\underline{x}^{(j)}) \\ \phi_2(\underline{x}^{(j)}) \\ \phi_3(\underline{x}^{(j)}) \end{bmatrix} = \begin{bmatrix} 1 \\ x_2^{(j)} \\ x_3^{(j)} \end{bmatrix}$$

Unknown weights $\underline{w}$

$$\underline{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

The model becomes

$$t_j = \begin{bmatrix} 1 & x_2^{(j)} & x_3^{(j)} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} + e^{(j)}$$

$$\boxed{t_j = w_1 + x_2^{(j)} w_2 + x_3^{(j)} w_3 + \epsilon^{(j)}}$$

Combining the N observations

$$
\begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ \vdots \\ t_N \end{bmatrix}
=
\begin{bmatrix}
1 & x_1^{(1)} & x_2^{(1)} \\
1 & x_1^{(2)} & x_2^{(2)} \\
1 & x_1^{(3)} & x_2^{(2)} \\
\vdots & \vdots & \vdots \\
1 & x_1^{(N)} & x_2^{(N)}
\end{bmatrix}
\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}
+
\begin{bmatrix} \epsilon^1 \\ \epsilon^2 \\ \epsilon^3 \\ \vdots \\ \epsilon^N \end{bmatrix}
$$

$$\boxed{\underline{t} = \Phi \, \underline{w} + \underline{\epsilon}}$$

$$\overbrace{-[\underline{t} - \Phi \underline{w}]^T [\underline{t} - \Phi \underline{w}]}^{\|\underline{t} - \Phi \underline{w}\|^2}$$

$$p(\underline{t} \mid \underline{X}, \underline{w}) = \frac{e^{\frac{-[\underline{t} - \Phi \underline{w}]^T [\underline{t} - \Phi \underline{w}]}{2\sigma^2}}}{(2\pi\sigma^2)^{N/2}}$$

Log likelyhood

$$\ln[p(\underline{t} \mid \underline{X}, \underline{w})]$$

find $\underline{w}$ that maximized the log likelyhood of $\underline{t}$

MAP $\quad$ MLE

$p(\underline{w}|\underline{t})$ $\quad$ $p(\underline{t}|\underline{w})$

Maximize

$$\ln\left(p\left(\underline{t}\,|\,\underline{X},\underline{w}\right)\right) =$$

$$= -\frac{N}{2}\ln\left[2\pi\right] - \frac{N}{2}\ln(\sigma^2) - \frac{\left[\underline{t}-\Phi\underline{w}\right]^T\left[\underline{t}-\Phi\underline{w}\right]}{2\sigma^2}$$

or minimize the $-\ln\left(p\left(\underline{t}\,|\,\underline{X},\underline{w}\right)\right) \equiv F$

$$F = \frac{N}{2}\ln\left[2\pi\right] + \frac{N}{2}\ln(\sigma^2) + \frac{\left[\underline{t}-\Phi\underline{w}\right]^T\left[\underline{t}-\Phi\underline{w}\right]}{2\sigma^2}$$

1° $\quad \frac{\partial F}{\partial \underline{w}} = 0 \implies \phi^T\left[\underline{t}-\Phi\underline{w}\right] = 0$

soln:

$$\boxed{\underline{w}_{mle} = \left[\phi^T\phi\right]^{-1}\Phi\,\underline{t}}$$

2° $\quad \frac{\partial F}{\partial \sigma^2} = \frac{N}{2}\frac{1}{\sigma^2} + \left[\underline{t}-\Phi\underline{w}\right]^T\left[\underline{t}-\Phi\underline{w}\right]\left\{\frac{-1}{2(\sigma^2)^2}\right\} = 0$

$$\boxed{\sigma^2 = \frac{1}{N}\left[\underline{t}-\Phi\underline{w}\right]^T\left[\underline{t}-\phi\underline{w}\right]\Big|_{\underline{w}=\underline{w}_{mle}}}$$

proof $\underline{w}_{mle}$

$$\frac{\partial F}{\partial \underline{w}} = \frac{1}{2\sigma^2} \frac{\partial}{\partial \underline{w}} \left[\left(\underline{t}^T - \underline{w}^T \phi^T\right) \left(\underline{t} - \phi \underline{w}\right) \right]$$

$$= \frac{1}{2\sigma^2} \frac{\partial}{\partial \underline{w}} \left[\underline{t}^T \underline{t} - \underline{t}^T \phi \underline{w} - \underline{w}^T \phi^T \underline{t} + \underline{w}^T \phi^T \phi \underline{w} \right]$$

note $\frac{\partial}{\partial \underline{a}} \left(\underline{a}^T \phi \underline{b} \right) = \phi \underline{b}$

$\frac{\partial}{\partial \underline{b}} \left(\underline{a}^T \phi \underline{b} \right) = \left(\underline{a}^T \phi \right)^T = \phi^T \underline{a}$

$$= \frac{1}{2\sigma^2} \left[- \phi^T \underline{t} - \phi^T \underline{t} + \phi^T \phi \underline{w} + \phi^T \phi \underline{w} \right]$$

$$= \frac{1}{2\sigma^2} \left[-2 \cdot \phi^T \underline{t} + 2 \phi^T \phi \underline{w} \right] = 0$$

$$\phi^T \phi \underline{w} = \phi^T \underline{t}$$

$$\boxed{\underline{w} = \left(\phi^T \phi \right)^{-1} \phi^T \underline{t}}$$

Max a Posterior prob (MAP)

Bayes thm

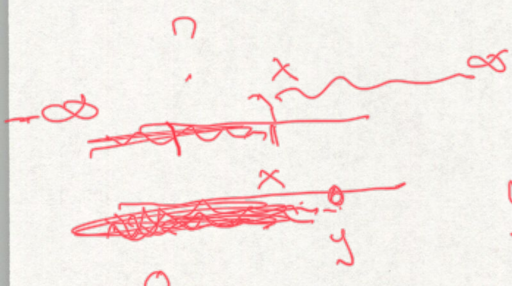
$$p(\underline{w}|\underline{t}) = \frac{p(\underline{t}|\underline{w})\,p(\underline{w})}{p(\underline{t})}$$

1° prior $\quad p(\underline{w}) = N_{\underline{w}}(0, I\alpha^2)$

2° Likelyhood

$$p(\underline{t}|\underline{w}) = \prod_{n=1}^{N} \hat{N}(t_n|\phi^T\underline{w})$$

$$p(\underline{w}|\underline{t}) = \frac{e^{-\frac{\|\underline{t}-\Phi\underline{w}\|^2}{2\sigma^2}}}{(2\pi\sigma^2)^{N/2}} \; \frac{e^{-\frac{\underline{w}^T\underline{w}}{2\alpha^2}}}{(2\pi\alpha^2)^{M/2}} \; \frac{1}{p(\underline{t})}$$



$y = ax + \epsilon \qquad a < 0$

$y = ax + \epsilon$

$$\boxed{\frac{y-\epsilon}{a} = \underline{X}}$$

$-\infty \quad = -\infty$

$\mathrm{Prob}(\underline{X} \leq y) = \mathrm{Prob}(a\underline{X} + \epsilon \leq y)$

$\mathrm{Prob}(a\underline{X} \leq y - \epsilon)$

$\mathrm{Prob}\left(\underline{X} \leq \frac{y-\epsilon}{a}\right)$

$$\boxed{\mathrm{Prob}\left(\underline{X} \geq \frac{y-\epsilon}{a}\right)}$$

$1 - \mathrm{Prob}\left(\underline{X} \leq \frac{y-\epsilon}{a}\right)$

MAP (estimate)

$$\ln\left[p(\underline{w}\,|\,\underline{t})\right] = \ln\left(p(\underline{t}\,|\,\underline{w})\right) + \ln\left(p(\underline{w})\right) - \ln\left(p(\underline{t})\right)$$

$$F = -\ln\left(p(\underline{t})\right) - \frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma^2) - \frac{M}{2}\ln(\alpha^2) - \frac{M}{2}\ln(2\pi)$$

$$- \frac{\|\underline{t} - \Phi\underline{w}\|^2}{2\sigma^2} - \frac{\underline{w}^T\underline{w}}{2\alpha^2}$$

maximize F

$$\frac{\partial F}{\partial \underline{W}} = 0 = -\left[\frac{\Phi^T\underline{t} - \Phi^T\Phi\underline{w}}{\sigma^2}\right] - \frac{I\,\underline{W}}{\alpha^2} = 0$$

$$\boxed{\left[\frac{\Phi^T\Phi}{\sigma^2} + \frac{I}{\alpha^2}\right]\underline{w} = \frac{\Phi\,\underline{t}}{\sigma^2}}$$

$$\frac{\partial F}{\partial \sigma^2} = 0 = -\frac{N}{2}\frac{1}{\sigma^2} + \frac{\|\underline{t} - \Phi\underline{w}\|^2}{2(\sigma^2)^2} = 0$$

$$\boxed{\sigma^2 = \frac{1}{N}\|\underline{t} - \Phi\underline{w}\|^2}$$

$$\boxed{\underline{w}_{MAP} = \left[\Phi^T\Phi + \frac{I\sigma^2}{\alpha^2}\right]^{-1}\left[\Phi^T\underline{t}\right]}$$

$(I\beta^2)$

$\underline{w} \mid J$ — MLE
max
abs — MAP
$\underline{w}$

Parametric distribution of $\underline{w}$ $\{ p(\underline{w} \mid \underline{t}) \}$

$$p(\underline{w} \mid \underline{t})\, p(\underline{t}) = p(\underline{t} \mid \underline{w})\, p(\underline{w})$$

$\underline{m}_0$

Σ_0^{-1}

① prior $p(\underline{w})$

$$p(\underline{w}) = \frac{e^{-[\underline{w}-\underline{m}_0]^T \Sigma_0^{-1} [\underline{w}-\underline{m}_0] \frac{1}{2}}}{(2\pi)^{M/2} \left| \Sigma_0 \right|^{1/2}}$$

where $\underline{w} = \begin{Bmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{Bmatrix}$; $\underline{m}_0$ mean ; Σ_0 covariance of $\underline{w}$

② likelyhood

$$p(\underline{t} \mid \underline{w}) = \frac{e^{-\frac{\| \underline{t} - \Phi \underline{w} \|^2}{2\sigma^2}}}{(2\pi \cdot \sigma^2)^{N/2}}$$

③ $$p(\underline{w} \mid t)\, p(\underline{t}) \propto \frac{e^{-[\underline{w}-\underline{m}_N]^T \Sigma_N^{-1} [\underline{w}-\underline{m}_N] \frac{1}{2}}}{(2\pi)^{M/2} \left| \Sigma_N \right|^{1/2}}$$

estimate $\underline{m}_N$, Σ_N

model

$\underline{m}_N$, Σ_N

Momentum Matching to find

$\underline{m}_N$ & Σ_N

$$\frac{1}{2}\left[\underline{w}-\underline{m}_0\right]^T \Sigma_0^{-1}\left[\underline{w}-\underline{m}_0\right]+\left[\underline{t}-\Phi\underline{w}\right]^T\left[\underline{t}-\phi\underline{w}\right]\frac{1}{2\sigma^2}$$

$$=\left[\underline{w}-\underline{m}_N\right]^T\Sigma_N^{-1}\left[\underline{w}-m_N\right]\frac{1}{2}$$

Σ_N

$$\frac{1}{\sigma^2}\left[w^T\phi^T\phi w+t^Tt-w^T\phi^Tt-t^T\phi\underline{w}\right]$$

$$+\left[w^T\Sigma_0^{-1}w+m_0^T\Sigma_0^{-1}m_0-w^T\Sigma_0^{-1}m_0-m_0^T\Sigma_0^{-1}w\right]$$

$$=\left[w^T\Sigma_N^{-1}w+m_N^T\Sigma_N^{-1}m_N-m_N^T\Sigma_N^{-1}\underline{w}-\underline{w}^T\Sigma_N^{-1}\underline{m}_N\right]$$

(1) $$\frac{w^T\phi^T\phi w}{\sigma^2}+w^T\Sigma_0^{-1}w=w^T\Sigma_N^{-1}w$$

$$\boxed{\frac{\phi^T\phi}{\sigma^2}+\Sigma_0^{-1}=\Sigma_N^{-1}}$$

(2) $$\frac{t^T\phi\underline{w}}{\sigma^2}+m_0^T\Sigma_0^{-1}\underline{w}=m_N^T\Sigma_N^{-1}\underline{w}$$

$$\frac{t^T\phi}{\sigma^2}+m_0^T\Sigma_0^{-1}=m_N^T\Sigma_N^{-1}$$

$$\left(\frac{t^T \phi}{\sigma^2} + m_0^T \Sigma_0^{-1} \right)^T = \left(m_N^T \Sigma_N^{-1} \right)^T$$

note $\left(\Sigma_N^{-1}\right)^T = \Sigma_N^{-1}$ $\quad \Sigma_N^{-1}$ is symmetric

likewise Σ_0^{-1} is sym

$$\left[\frac{\phi^T t}{\sigma^2} + \Sigma_0^{-1} m_0 \right] = \Sigma_N^{-1} m_N$$

1

$$\boxed{m_N = \Sigma_N \left[\frac{\phi^T t}{\sigma^2} + \Sigma_0^{-1} m_0 \right]}$$