$$L(\omega) = \frac{1}{2}\sum \frac{1}{2}(y-T)^2$$
$$J(\omega) = \frac{1}{2}\sum \frac{1}{2}(y-T)^2 + \frac{\alpha}{2}\|\omega\|^2$$
$y = XW$

$$\nabla L(\omega) = X^T(XW - T) = 0$$
$$= (X^TX)\omega - X^TT = 0$$

$$\nabla J(\omega) = X^T(XW-T) + \alpha\omega$$
$$= X^TXW + \alpha W - X^TT$$
$$= (\underbrace{X^TX}_{A} + \underbrace{\alpha I}_{\alpha I})\omega = X^TT$$
$$= \underbrace{X^TT}_{B} \quad \text{(Norm eqn)}$$

Ridge/LASSO → sparse representation

$$F(x) = Ax^2 + Bx + C$$
$$\hookrightarrow G(x) = \underline{\underline{A}}x^3 + Bx^2 + Cx + D$$
$$\downarrow 0$$

$L(\omega)$ wrt $\|W\|^2 \leq 1$

$$L(\omega) + \lambda(\|W\|^2 - 1)$$

→ ω^* is optimal

$$L(\omega) = L(\omega^*) + \cancel{\nabla L(\omega^*)(\omega - \omega^*)}$$
$$+ \frac{1}{2}(\omega - \omega^*)\nabla\nabla L(\omega^*)(\omega - \omega^*)$$

$$= L(\omega^*) + \frac{1}{2}(\omega - \omega^*)\underbrace{\nabla\nabla L(\omega^*)}_{H}(\omega - \omega^*)$$

$$L(\omega) = L(\omega^*) + \frac{1}{2}(\omega - \omega^*) H (\omega - \omega^*)$$

$$\nabla(L(\omega)) = H(\omega - \omega)^*$$

$$J(\omega) = L(\omega) + \frac{\alpha}{2}\|\omega\|^2$$

$$\nabla J(\omega) = \nabla L(\omega) + \alpha \omega$$

$$= H(\omega - \omega^*) + \alpha\omega$$

$\tilde{\omega} \Rightarrow$ optimal solution for $J(\omega)$ &

ω^* for $L(\omega)$

$\nabla J(\tilde{\omega}) = 0$

$\Rightarrow H(\tilde{\omega} - \omega^*) + \alpha\tilde{\omega}$

$$(H + \alpha I)\tilde{\omega} = H\omega^*$$

$$\tilde{\omega} = (H + \alpha I)^{-1} H\omega^*$$

Norm:

$$\omega = (X^T X + \alpha I)^{-1} X^T T$$

$\boxed{V^T H V \geq 0 \quad \forall V}$ semidefinite

$\boxed{V^T H V > 0 \quad \forall V \quad H \text{ is +ve definite}}$

$\langle V^T | H | V \rangle \geq 0$

$H = Q \Lambda Q^T$ (Eigenvalue decomp)

$\Lambda \rightarrow$ eigenvalue

$Q \Rightarrow$ eigenvectors

$\downarrow$

basis

$$QQ^T = I = Q^T Q$$

Diagonal Matrix: $\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$

inverse: $\begin{bmatrix} 1/a & 0 & 0 \\ 0 & 1/b & 0 \\ 0 & 0 & 1/c \end{bmatrix}$

$$\omega_{t+1} = \omega_t - \eta \nabla J(\omega)$$

$$\tilde{\omega} = (H + \alpha I)^{-1} H\omega^* \qquad H = Q\Lambda Q^T$$

$$= (Q\Lambda Q^T + \alpha \cdot Q I Q^T)^{-1} Q\Lambda Q^T \omega^*$$

$$(\because QQ^T = I)$$

$$= (Q(\Lambda+\alpha I)Q^T)^{-1} Q\Lambda Q^T w^*$$

$$(Q^T)^{-1}(\Lambda+\alpha I)\cancel{Q^{-1}Q}\Lambda Q^T w^*$$

$$Q \underbrace{(\Lambda+\alpha I)^{-1}\Lambda}_{\text{Diagonal}} Q^T w^*$$

$$\Lambda,\ \Lambda+\alpha I,\ (\Lambda+\alpha I)^{-1},\ (\Lambda+\alpha I)^{-1}\Lambda$$

$$\lambda_i,\ \lambda_i+\alpha,\ \frac{1}{\lambda_i+\alpha},\ \frac{\lambda_i}{\lambda_i+\alpha}$$

↳ i^{th} element

$$\tilde{w} = QDQ^T w^*$$

↓

$\frac{\lambda_i}{\lambda_i+\alpha} \Rightarrow$ Scaling

matrix $MM^{-1} \Rightarrow$ if M rotates MM^{-1} brings back

So youre rotating, scale, rotated

$\Rightarrow \tilde{w} \propto k w^*$ Scale down or stay

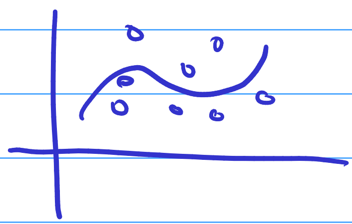
Scaling factor in accordance to eigenvalue

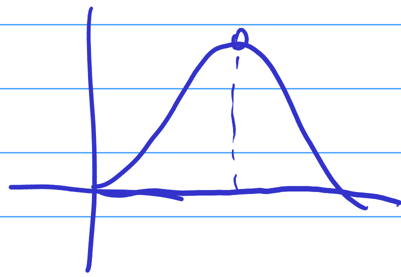
For large $\lambda \Rightarrow$ more prominent
w stays

For smaller $\lambda \Rightarrow$ less prominent
w dies

$$\sum \lambda_i = \text{trace}(A)$$

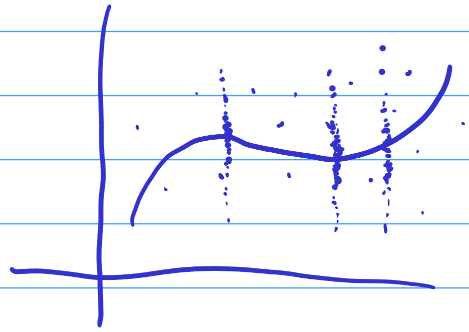
$$\prod \lambda_i = \det(A)$$





$P(x,y)$

$$P(y) = \int_y P(x,y)\,dy$$



predicted function

normal

iid → MLE $P(x_i\ x_j\ x_k \cdots)$

$$(2\pi\sigma^2)^{-N} \exp\left\{-\frac{(t_i - y_i)^2}{2\sigma^2}\right\} = M$$

$$\log M = -N\log(2\pi\sigma^2) - \left(\frac{t_i - y_i}{2\sigma^2}\right)^2$$

$$y_i = x_i w$$

derivative ⇒ $\frac{\cancel{2}(t_i - x_i w)}{\cancel{2}\sigma^2} x_i = 0$

$$\Rightarrow (t_i - x_i w)\,x_i$$

$$\max -(t_i - x_i w)^2$$

$$\Rightarrow \min\left((t_i - x_i w)^2\right)$$

 → l = uniform 76–78 cm

given few observations
find new mean!

$P(\mu) \rightarrow$ prior, uniform distribution

$P(obs) \Rightarrow$ normal

$$P(\mu) \underbrace{\pi P(obs)}_{MLE} \Rightarrow \text{Posterior}$$

Prior $\Rightarrow$ normal distribution around 0

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(t_i - y_i)^2}{2\sigma^2}}$$

$$P(\omega/t) \propto P(\omega) P(t/\omega)$$

$$P(\omega/t, x) \propto P(\omega) \prod_{i=1}^{N} P(t_i/\omega x_i)$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{\frac{-\omega^2}{2\sigma_1^2}} \left[\frac{1}{\sqrt{2\pi\sigma_2^2}}\right]^N e^{-\sum\frac{(y_i - t_i)^2}{2\sigma^2}}$$

$$\log L = -\frac{1}{2}\log(\cancel{2\pi}\sigma_1^2) - \frac{\|\omega\|^2}{2\sigma_1^2} - \frac{N}{2}\log(\cancel{2\pi}\sigma_2^2) - \sum\frac{(y_i - t_i)^2}{2\sigma_2^2}$$

$$\max L = -\sum\frac{(y_i - t_i)^2}{2\sigma_2^2} - \frac{\|\omega\|^2}{2\sigma_1^2}$$

$$\min L = \sum_{i=1}^{N}(y_i - t_i)^2 + \underbrace{\left(\frac{\sigma_2}{\sigma_1}\right)^2}_{\rightarrow \lambda} \|\omega\|^2$$

$= $ Regularized loss

Normal