Machine Learning

Supervised learning

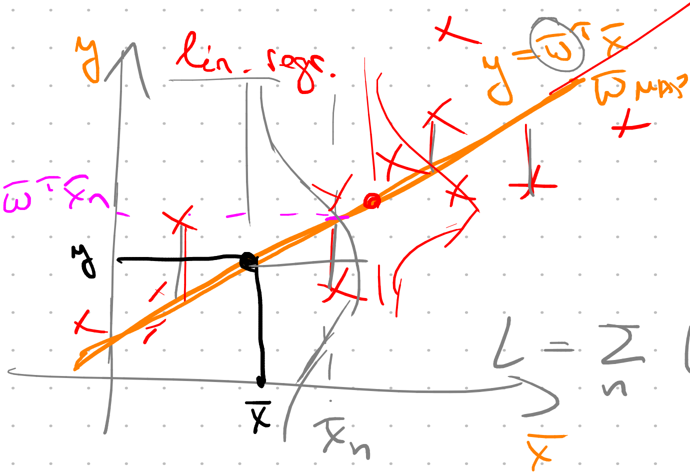
$D = \{(\bar{x}_n, y_n)\}_{n=1}^{N}$

Regression — Classification

Learning to rank

$p(y|\bar{x})$

$\bar{x}$ — [ ] — $p(y|\bar{x})$



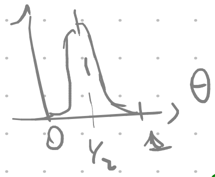
Unsupervised Learning

$D = \{\bar{x}_n\}_{n=1}^{N}$

$p(\bar{x})$

Reinforcement Learning

Agent ⇄ Env



dataset, model parameters, prior distr.

posterior distrib.

$$p(\bar{\theta}|D) = \frac{p(D|\bar{\theta})\,p(\bar{\theta})}{p(D)}$$

likelihood

$$p(D|\bar{\theta}) = \prod_{n=1}^{N} p(\bar{x}_n|\bar{\theta})$$

$\bar{\omega}$

$$p(D|\bar{\theta}) = p(\bar{y}|\bar{\theta}, X) = \prod_{n=1}^{N} p(y_n|\bar{\theta}, \bar{x}_n)$$

$$L = \sum_n (y_n - \bar{\omega}^T\bar{x}_n)^2$$

$$y = \bar{\omega}^T\bar{x} + \varepsilon \sim \mathcal{N}(\varepsilon|0, \sigma^2)$$

1 3 2 4 μ

$$p(D|\bar{\omega}) = \prod_n \mathcal{N}(y_n|\bar{\omega}^T\bar{x}_n, \sigma^2) = \prod_n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_n - \bar{\omega}^T\bar{x}_n)^2} \xrightarrow[\bar{\omega}]{} \max$$

$$\log p(D|\bar{\omega}) = -\frac{N}{2}\log(2\pi\sigma^2) - \sum_{n=1}^{N} \frac{1}{2\sigma^2}(y_n - \bar{\omega}^T\bar{x}_n)^2$$

$$\underbrace{\sum (y_n - \bar{\omega}^T\bar{x}_n)^2}_{\bar{\omega}\ \to\ \min}$$

$e^{-(\theta - \frac{1}{2})^2}$

$\theta = p(\text{орёл})$, $D = \text{htthhht}$

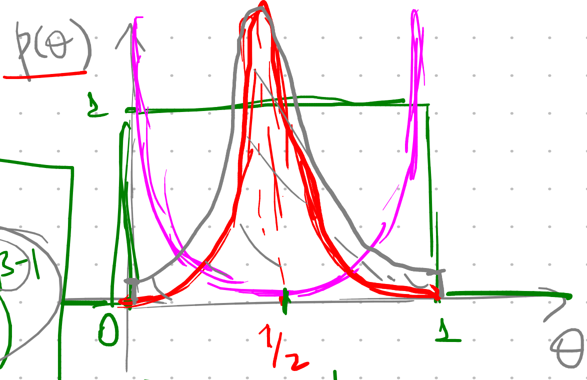
$$p(D|\theta) = \prod_n p(d_n|\theta) = \theta^{n}(1-\theta)^{m} \xrightarrow[\theta]{} \max \qquad 0, 1 \qquad \theta_{ML} = \frac{n}{n+m}$$

$p(D|\theta) = \theta^n \to \infty$

$$n\cdot\theta^{n-1}(1-\theta)^m - m\cdot\theta^n(1-\theta)^{m-1} = 0; \quad \theta^{n-1}(1-\theta)^{m-1}\left(n(1-\theta) - m\theta\right) = 0$$

$p(\theta)$ – prior

$p(\theta|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$, $\theta \in [0,1]$



Conjugate prior

$\bar\theta \to \max$

$$p(\bar\theta|D) \propto \left[\propto p(\bar\theta)\, p(D|\bar\theta)\right]$$

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

$$\log p(\bar\theta|D) = \text{Const} + \log p(D|\bar\theta) + \log p(\bar\theta) \xrightarrow{\bar\theta} \max$$

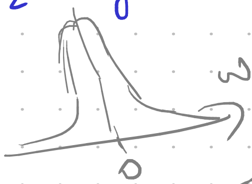
regularizer

LR:

$$-\left[\frac{1}{2\sigma^2}\sum^{N}(y_n - \bar w^T \bar x_n)^2 + R(\bar\theta)\right] \xrightarrow{\theta} \min$$

σ^2/σ_0^2

L_2-regularization:



$$\sum(y_n - \bar w^T \bar x_n)^2 + \alpha \cdot \|w\|^2 \longrightarrow \min$$

$-c \cdot \log p(\bar w)$

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

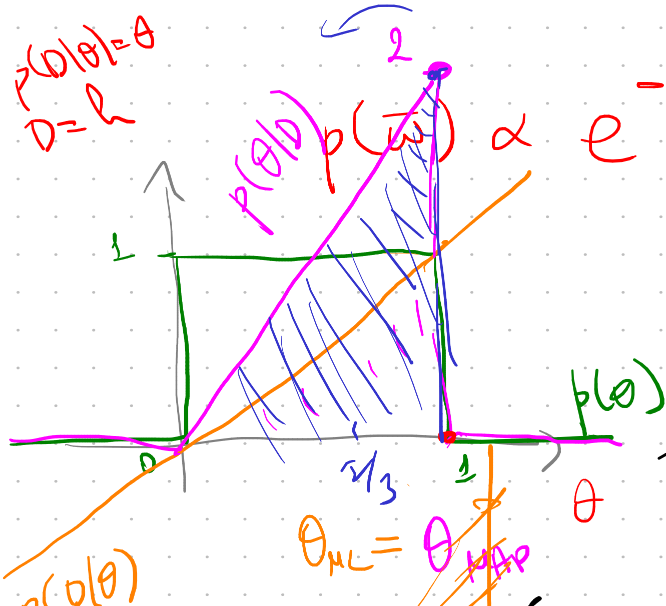
L_1-reg.

$$\sum(\ \ )^2 + \alpha \cdot \sum_i |w_i|$$

Lasso regression

$p(D|\theta) = \theta$
$D = h$

$$p(\bar w) \propto e^{-c \cdot \sum_{i=1}^{d} |w_i|}$$



$$p(y|D, \bar x)$$

$$\boxed{p(\text{open}|D)} = \int_\theta p(\text{open}|\theta) \cdot p(\theta|D)\, d\theta = \int \theta \cdot \text{Const} \cdot \theta^{n+\alpha-1}(1-\theta)^{n+\beta-1} d\theta = \dots$$

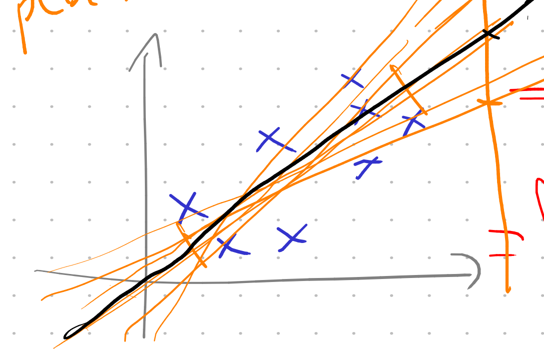
predictive distribution

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

$$= \int p(\bar x|\bar\theta, \cancel{D})\, p(\bar\theta|D)\, d\bar\theta$$

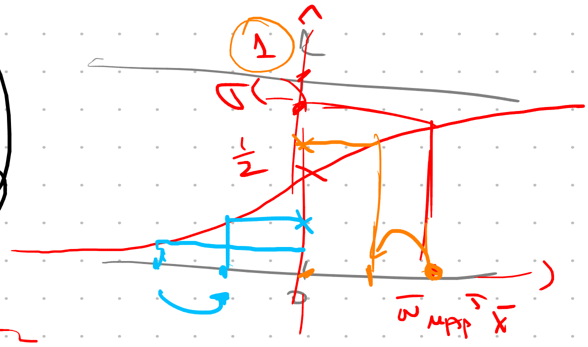
likelihood of $\bar x$ — posterior

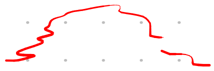
$$= \int p(\bar x|\bar\theta)\, p(\bar\theta|D)\, d\bar\theta = E_{p(\theta|D)}\left[p(\bar x|\bar\theta)\right]$$

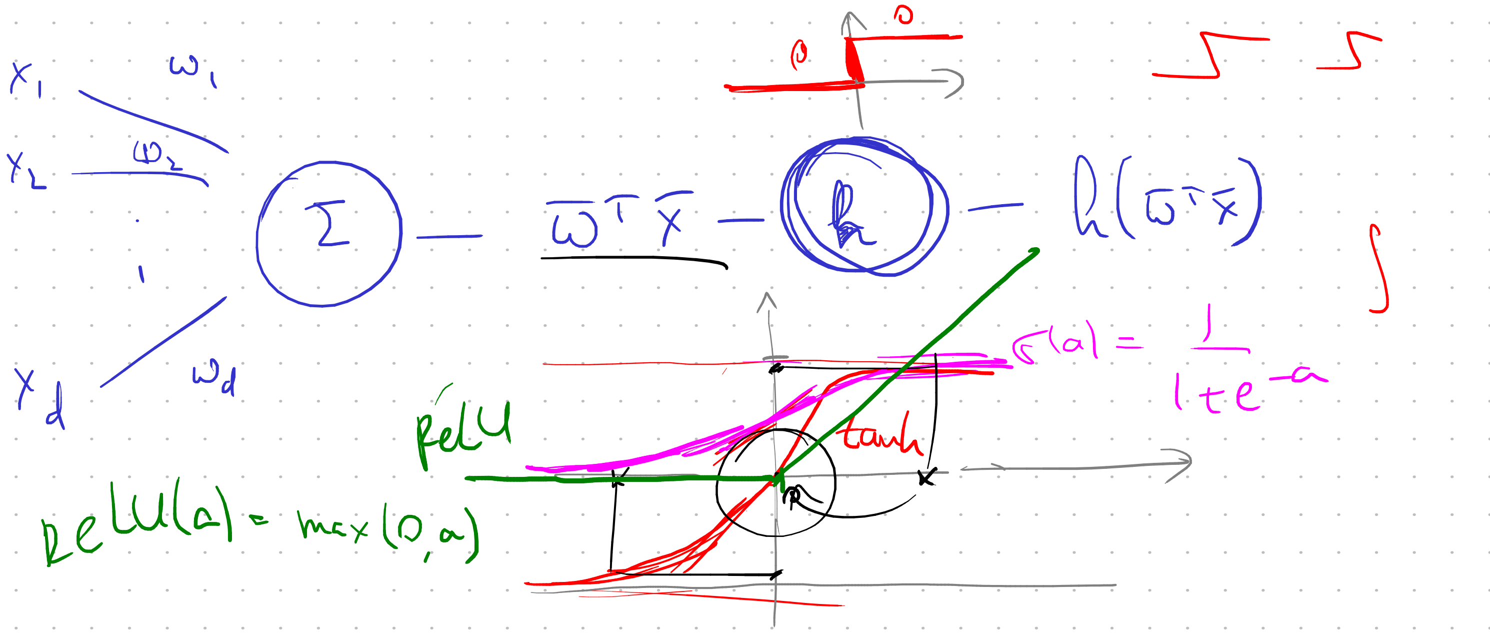


$$p(y|\bar{x}, D) = \mathcal{N}\left(y \,\middle|\, \bar{\omega}_{MAP}^T \bar{x},\ \sigma^2 + \underbrace{\ldots(\bar{x})}\right)$$

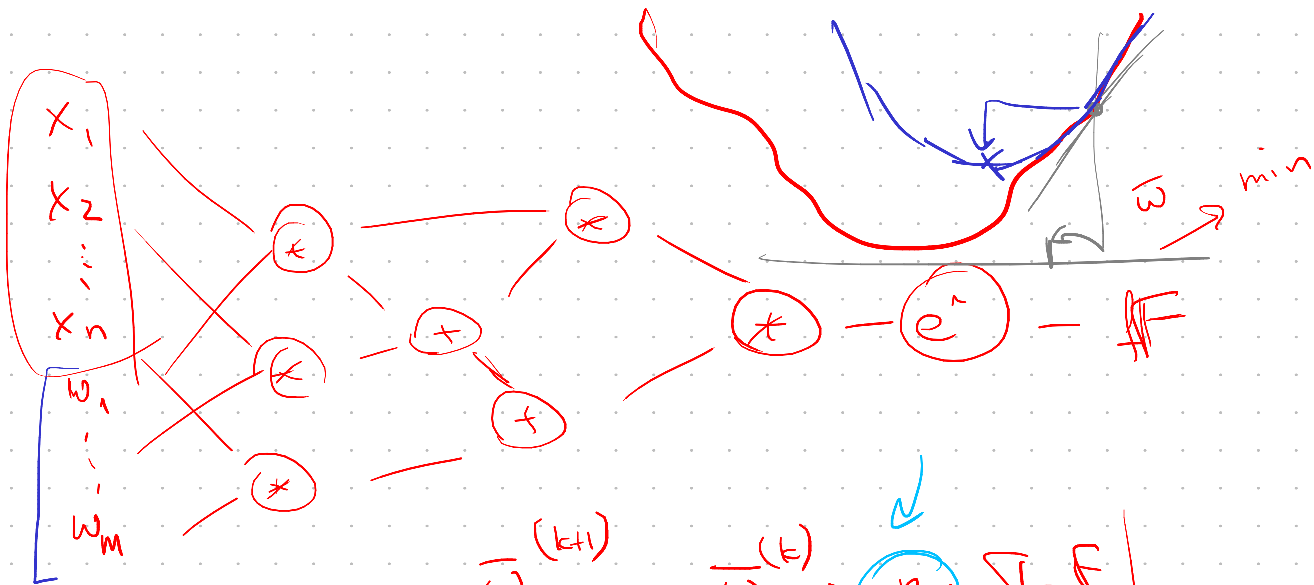
$$p(C_1|\bar{x}, D) \approx \sigma\left(\frac{\bar{\omega}_{MAP}^T \bar{x}}{\underbrace{1 + \kappa^2(\bar{x})}}\right)$$







$reLU(a) = \max(0, a)$



$$\bar{\omega}^{(k+1)} = \bar{\omega}^{(k)} - \eta \nabla_{\bar{\omega}} F \Big|_{\bar{\omega}^{(k)}}$$

$$F(\bar{\omega}) = F(\bar{\omega}^{(k)}) + \nabla_{\bar{\omega}} F(\bar{\omega}^{(k)})^T (\bar{\omega} - \bar{\omega}^{(k)}) + \frac{1}{2}(\bar{\omega} - \bar{\omega}^{(k)})^T H (\bar{\omega} - \bar{\omega}^{(k)}) + \ldots$$

$$\boxed{\bar{\omega}_*^{(k)} = \bar{\omega}^{(k)} - H^{-1} \nabla_{\bar{\omega}} F(\bar{\omega}^{(k)})}$$

Hessian $\left(\frac{\partial F}{\partial \omega_i \partial \omega_j}\right)$

quasi-Newton methods $\bar{\omega}$, $\cancel{\nabla_{\bar{\omega}} F} \rightsquigarrow \hat{H}^{-1} \approx H$ $\boxed{l\text{-BFGS}}$

$$\nabla_{\bar\omega} L = \frac{1}{N}\sum_{n=1}^{N} \nabla_{\bar\omega}\, l(\bar\omega, \bar x_n) = \mathbb{E}_{\underline{\text{unif}(D)}}\left[\nabla_{\bar\omega}\, l(\bar\omega,\bar x)\right]$$

Stochastic gradient descent:

$$F(\bar x) = \mathbb{E}_{q(\bar y)}\left[f(\bar x,\bar y)\right]$$

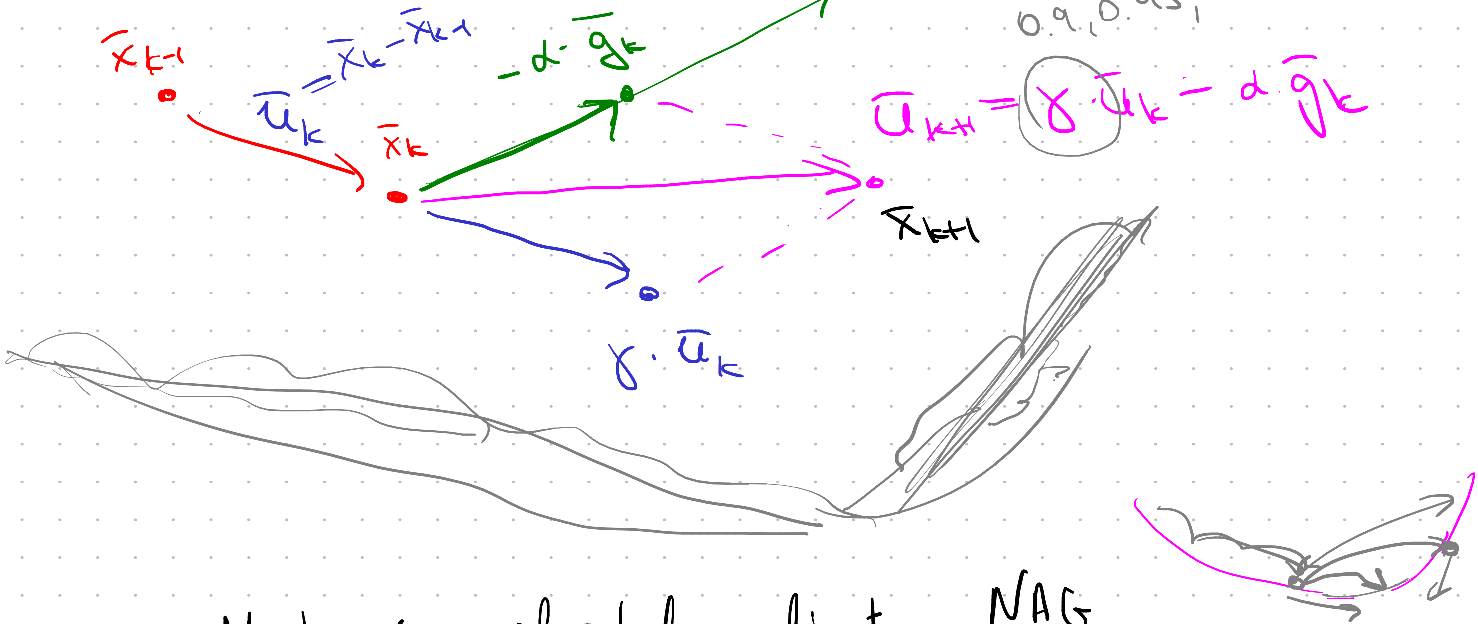
$$\hat F(\bar x^{(t)}) \simeq \frac{1}{R}\sum_r f(\bar x^{(t)}, \bar y^{(r)}), \quad \hat G(\bar x^{(t)}) = \frac{1}{R}\sum - \approx \nabla_{\bar x} F$$

Momentum

$$-\nabla_{\bar x} F|_{\bar x_k} = \bar g_k$$

0.9, 0.95, 0.995

$$\bar u_{k+1} = \gamma\cdot\bar u_k - \alpha\cdot\bar g_k$$



Nesterov accelerated gradient NAG

$$\bar x_{k+1} = \bar x_k + \gamma\cdot\bar u_k - \alpha\nabla_{\bar x} F|_{\bar x_k+\gamma\cdot\bar u_k}$$