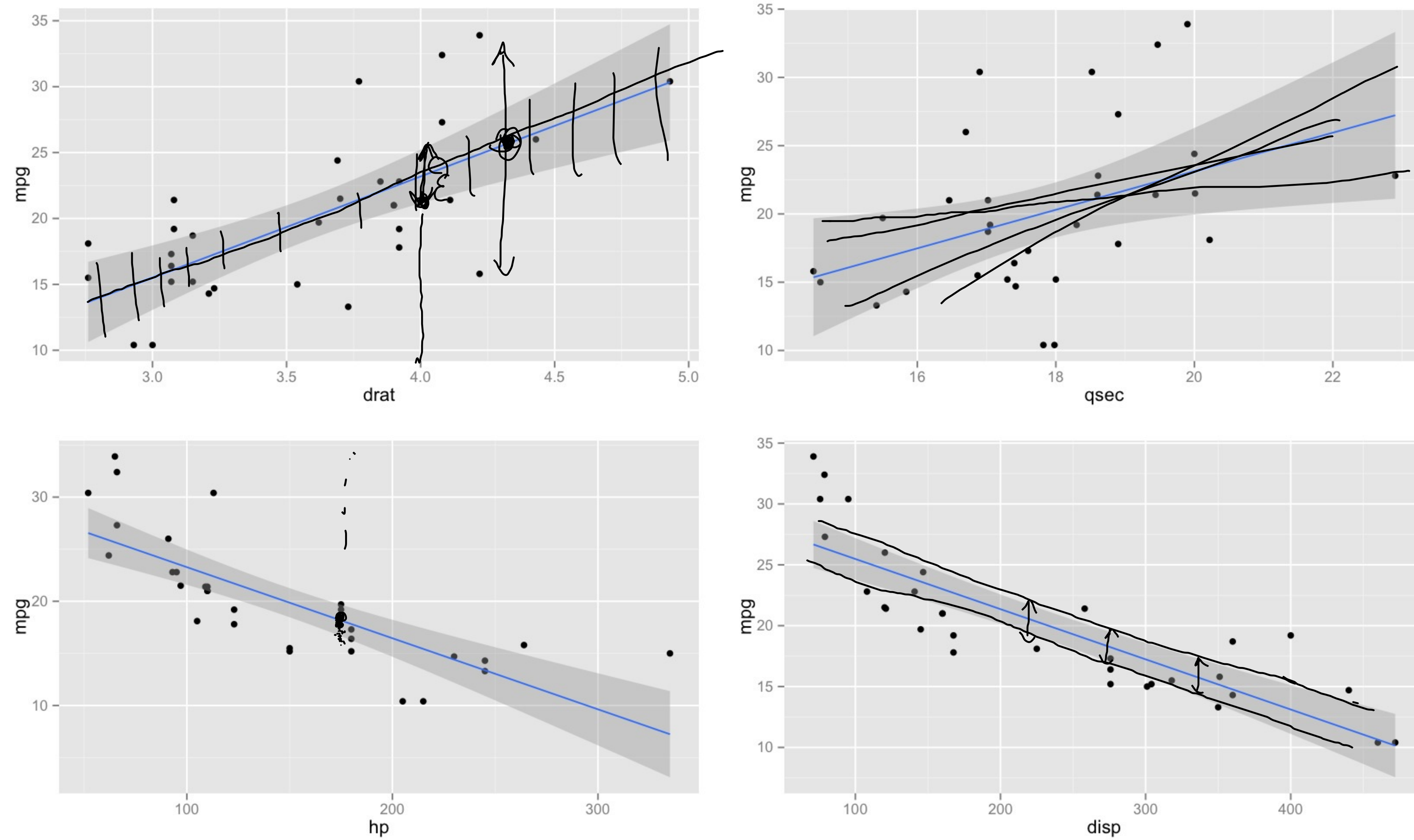
$x \quad y$

$$y \approx w^T x = w_0 + w_1 x_1 + \dots + w_d x_d$$

$$y = w^T x + \varepsilon \quad \rightarrow \text{assume} \quad \varepsilon \sim N(0, \sigma^2)$$



95%.

$(x^{(i)}, y^{(i)})$

$y \cong w^T x + \varepsilon$

$\Rightarrow \varepsilon \cong w^T x - y$

$\sigma^2 = \mathrm{Var}(\varepsilon)$

$\varepsilon \sim N(0, \sigma^2)$

$\mathbb{E}\varepsilon$

$\Pr[0 \le \varepsilon \le 1 \cdot \sigma] \simeq 34.1\%$

$\Pr[-\sigma \le \varepsilon \le \sigma] \simeq 68.2\%$

$\Pr[-2\sigma \le \varepsilon \le 2\sigma] \simeq 95.4\%$

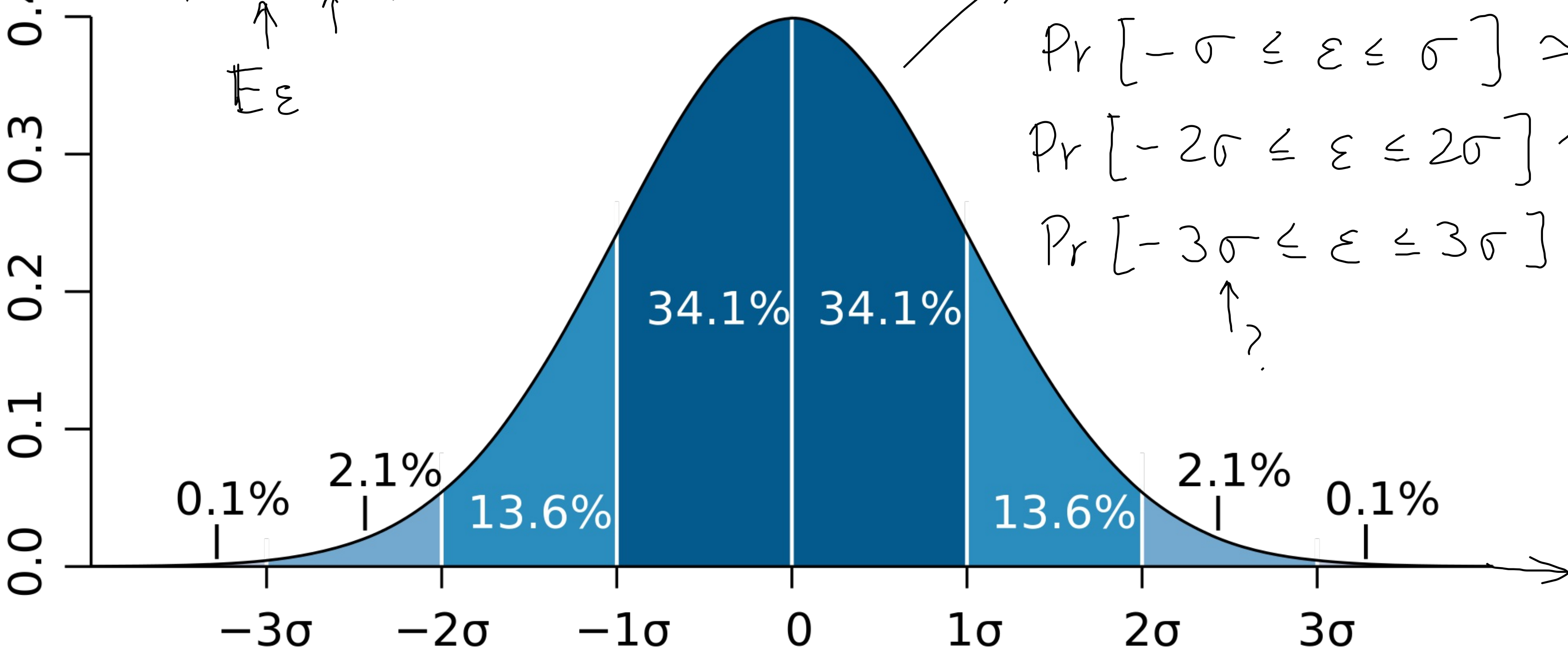
$\Pr[-3\sigma \le \varepsilon \le 3\sigma] \simeq 99.6\%$

?

$\hat{w}_0$

$\hat{w}_1$

$\vdots$



ε

$\pm 2\sigma$

$\sigma^2 = \mathrm{Var}(\varepsilon) = \mathbb{E}\left[(\varepsilon - \mathbb{E}\varepsilon)^2\right]$

$= \mathbb{E}[\varepsilon^2]$

$\hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m} (\varepsilon^{(i)})^2 = \frac{1}{m}\sum_{i=1}^{m} (w^T x^{(i)} - y^{(i)})^2$

MSE (Mean Squared Error) $\xrightarrow{\sqrt{\cdot}}$ Root MSE $\hat{\sigma}$

Confidenza per la classificazione generativa

Obiettivo: $\rightarrow \Pr(y=j \mid x) \Leftarrow \begin{cases} \Pr(y=j) = \pi_j \\ \Pr(x \mid y=j) = P_j(x) \end{cases}$

Esempio: $j \in \{0,1\}$

$$\underline{\Pr(y=1 \mid x)} = \frac{\Pr(x, y=1)}{\Pr(x)} = \frac{\Pr(x \mid y=1)\cdot\Pr(y=1)}{\Pr(x, y=1) + \Pr(x, y=0)} = \frac{\Pr(x \mid y=1)\,\pi_1}{\Pr(x \mid y=1)\,\pi_1 + \Pr(x \mid y=0)\,\pi_0}$$

0.8

0.51

$$= \frac{P_1(x)\,\pi_1}{P_1(x)\,\pi_1 + P_0(x)\,\pi_0} = \frac{1}{1 + \underbrace{\frac{P_0(x)\,\pi_0}{P_1(x)\,\pi_1}}_{=\exp(-a)}} = \frac{1}{1+\exp(-a)}$$

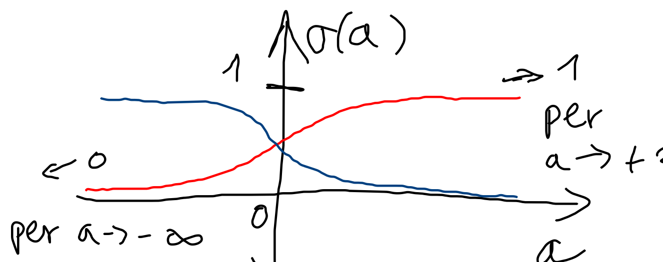
Funzione sigmoide $\sigma(a)$

$$\underline{\Pr(y=0 \mid x)} = \frac{P_0(x)\,\pi_0}{P_1(x)\,\pi_1 + P_0(x)\,\pi_0} = \frac{1}{1+\exp(+a)}$$

0.2

0.49

$$a = \log \frac{P_1(x)\,\pi_1}{P_0(x)\,\pi_0} = \underbrace{\log(P_1(x)\pi_1)}_{\delta_1(x)} - \underbrace{\log(P_0(x)\pi_0)}_{\delta_0(x)} = \delta_1(x) - \delta_0(x)$$



Maximum Likelihood Estimation di una distribuzione di Bernoulli

$y \in \{0,1\}$ $\quad \left.\begin{aligned} Pr(y=1) &= \pi \\ Pr(y=0) &= 1-\pi \end{aligned}\right\}$ $\quad Pr(y;\pi) = \pi^y (1-\pi)^{1-y} = \begin{cases} 1-\pi & \text{se } y=0 \\ \pi & \text{se } y=1 \end{cases}$

$(Pr(y))$

$$\mathcal{L}(\pi; y^{(1)}, \dots, y^{(m)}) = \prod_{i=1}^{m} Pr(y^{(i)};\pi) = \prod_{i=1}^{m} \pi^{y^{(i)}} (1-\pi)^{(1-y^{(i)})}$$

$$\operatorname{argmax}_\pi \mathcal{L}(\pi; y^{(1)}, \dots, y^{(m)}) = \operatorname{argmax}_\pi \log \mathcal{L}(\pi; \dots) = \operatorname{argmax}_\pi \sum_{i=1}^{m} \left[y^{(i)} \log \pi + (1-y^{(i)}) \log(1-\pi) \right]$$

$$= \operatorname{argmin}_\pi \underbrace{\sum_{i=1}^{m} \left[-y^{(i)} \log \pi - (1-y^{(i)}) \log(1-\pi) \right]}_{\substack{\text{Cross-entropia (convessa)} \\ g(\pi)}} \iff \nabla g(\pi) = 0$$

$$g'(\pi) = \sum_{i=1}^{m} \left[-\frac{y^{(i)}}{\pi} + \frac{(1-y^{(i)})}{1-\pi} \right] \overset{?}{=} 0$$

$$g'(\pi) = \sum_{i=1}^{m} \left[-\frac{y^{(i)}}{\pi} + \frac{(1-y^{(i)})}{1-\pi} \right] \stackrel{?}{=} 0$$

$$\iff \sum_{i=1}^{m} \frac{(1-y^{(i)})}{1-\pi} = \sum_{i=1}^{m} \frac{y^{(i)}}{\pi}$$

$$= \frac{1}{1-\pi} \sum_{i=1}^{m} (1-y^{(i)}) = \frac{1}{\pi} \underbrace{\sum_{i=1}^{m} y^{(i)}}$$

$$\frac{m - \sum_i y^{(i)}}{\sum_i y^{(i)}} = \frac{\sum_{i=1}^{m} (1-y^{(i)})}{\sum_{i=1}^{m} y^{(i)}} = \frac{1-\pi}{\pi} = \frac{1}{\pi} - \cancel{1}$$

$$= \frac{m}{\sum_i y^{(i)}} - \cancel{1}$$

$$\iff \hat{\pi} = \frac{\sum_{i=1}^{m} y^{(i)}}{m}$$