Bias-Variance decomposition

$$TSS = ESS + RSS$$

$$L(\mu) = E_{X,y}\left[E_x\left[(y - \mu(X)(x))^2\right]\right] =$$

$$= \underbrace{E_{x,y}\left[(\underbrace{y} - \underbrace{E[y|x]})^2\right]}_{\text{Шум}} + \underbrace{E_x\left[\left(E_X[\mu(X)(x)] - E[y|x]\right)^2\right]}_{\text{Смещение}} +$$

$$\cancel{E_x\left[\left(\underbrace{E_X[\mu(X)(x)]} - \underbrace{E[y|x]}\right)^2\right]}$$

~~Смещение~~

$$+ \underbrace{E_x\left[E_X\left[\left(\mu(X)(x) - E_X[\mu(X)(x)]\right)^2\right]\right]}_{\text{Разброс}}$$

Связь с регуляризацией

$\boxed{y = wx}$, $\quad X = \{(x_1, y_1), \ldots, (x_\ell, y_\ell)\}$

$$\boxed{L(w) = \frac{1}{\ell}\sum_{i=1}^{\ell}(y_i - wx_i)^2 + \lambda w^2}$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y = w(X)$$

$$\mu(X)(x) = w(X)\, x$$

$$w(X) = \frac{\sum x_i y_i}{\sum x_i^2 + \lambda \ell}$$

1) X – детерминирована

2) X – стохастические

Детерминированный X

$y = f(x) + \varepsilon$, ε – iid , $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2$

$y \sim (f(x), \sigma^2)$

Пусть истинная зависимость:

$$\boxed{f(x) = ax}$$

1) Шум. $E_{x,y}[(y - E[y|x])^2]$

$$E(y|x) = E(\underbrace{f(x) + \varepsilon}_{} | x) = f(x) \qquad Var(X) = E(X^2) - E(X)^2$$

$$E_{x,y}[(\cancel{f(x)} + \varepsilon - \cancel{f(x)})^2] = E_{x,y}[\varepsilon^2] =$$

$$= Var(\varepsilon) + \underset{\overset{\shortparallel}{0}}{E(\varepsilon)}^2 = \sigma^2$$

2) Смещение $E_x[(E_X[\mu(X)(x)] - E[y|x])^2]$

$$X = ((x_1, \overset{f(x)+\varepsilon}{y_1}), \dots (x_l, y_l))$$

$$E_X[\mu(x)(x)] = E_X[w(X)x] = E_X[w(X)](x)$$

$$E_X[w(X)] = E_X\left[\frac{\sum x_i y_i}{\sum x_i^2 + \lambda}\right] = E_{\cancel{X}\,\varepsilon}\left[\frac{\sum x_i(f(x_i) + \varepsilon_i)}{\sum x_i^2 + \lambda}\right] =$$

$$= E_\varepsilon\left[\frac{\sum x_i(ax_i+\varepsilon_i)}{\sum x_i^2+\lambda}\right] = E_\varepsilon\left[a\frac{\sum x_i^2}{\sum x_i^2+\lambda} + \frac{\sum x_i\varepsilon_i}{\sum x_i^2+\lambda}\right]$$

$$= a\frac{\sum x_i^2}{\sum x_i^2+\lambda} + \underbrace{\frac{\sum x_i E\varepsilon_i^{\to 0}}{\sum x_i^2+\lambda}}_{\sim 0}$$

$$E_x[\mu(X)(x)] - E[y|x])$$

$$\sum x_i^2 = s$$

$$= \left(a\frac{\sum x_i^2}{\sum x_i^2+\lambda}x - ax\right) = \left(a\frac{s}{s+\lambda}x - ax\right)$$

$$\frac{s}{s+\lambda} = 1 \Rightarrow \lambda = 0$$

$$E_x[(E_x[\mu(X)(x)] - E[y|x])^2] = \left(\frac{s}{s+\lambda}-1\right)^2 a^2 x^2$$

Разброс: $\underbrace{+\underbrace{E_x}\left[E_X\left[\left(\mu(X)(x) - E_X[\mu(X)(x)]\right)^2\right]\right]}_{\text{Разброс}}$

$$Var_X(\mu(X)(x)) = Var_X\left(\frac{\sum x_i(ax_i+\varepsilon_i)}{\sum x_i^2+\lambda}x\right) =$$

$$= \mathrm{Var}_X\left(\underbrace{a\frac{\sum x_i^2}{\sum x_i^2+\lambda}x}_{} + \frac{\sum x_i \varepsilon_i}{\sum x_i^2+\lambda}x \right) =$$

$$= \mathrm{Var}\left(\frac{\sum x_i \varepsilon_i}{\sum x_i^2+\lambda}x \right) = \frac{x^2}{(\sum x_i^2+\lambda)^2} \sum x_i^2 \underbrace{\mathrm{Var}(\varepsilon_i)}_{\sigma^2} =$$

$$\mathrm{Var}(\sum x_i \varepsilon_i) = \sum x_i^2 \mathrm{Var}(\varepsilon_i)$$

$$= \frac{s}{(s+\lambda)^2} x^2 \sigma^2$$

$$E_{x,y}\left[E_X\left[(y-\mu(X)(x))^2\right]\right] = \sigma^2 +$$

$$\underbrace{\left(\frac{s}{s+\lambda}-1\right)^2 a^2 x^2}_{0} + \underbrace{\frac{s}{(s+\lambda)^2} \underbrace{x^2\sigma^2}_{}}_{} \to \min_\lambda$$

$\lambda = 0$

$\lambda >> 0$

$$\gamma = \frac{1}{s+\lambda}$$

$$\sigma^2 + (s\gamma-1)^2 a^2 x^2 + s\gamma^2 x^2 \sigma^2 \to \min \lambda$$

$$(sa^2+\sigma^2)x^2 s\gamma^2 - 2sa^2x^2\gamma + a^2x^2 + \sigma^2$$

$$\gamma^* = \frac{1}{s+\sigma^2/a^2} \qquad \lambda^* = \frac{\sigma^2}{a^2} > 0$$

$\hat{\alpha} = \hat{\omega} =$

Деревья

1) Бутстрэп
2) Бэггинг

$y = \hat{\alpha} x$

Бутстрэп.

$X = ((x_1, y_1) \dots (x_l, y_l))$

$\hat{X}_1 = \dots \qquad \mu_1$

$\hat{X}_2 \dots \qquad \vdots$

$\vdots$

$\hat{X}_k \dots \qquad \mu_k$

Какие деревья строим?

m признаков

$\sqrt{m}$

При бинарных классификатора

$P(\text{ошибки}) = p$

$P(\text{ошибки голосования}) = f(p) \vee P$

При каких p вер-ть будет меньше p

$(1-p)^3$ – все три верно

$3p(1-p)^2$ – один ошибся

$$1-(1-p)^3-3p(1-p)^2 = p^2(3-2p)$$

$$-2p^3+3p^2<p$$

$$p(\underbrace{-2p^2+3p-1})<0$$

$$\underbrace{-2p}_{\le 0}\underbrace{(p-1)}_{\le 0}\underbrace{\left(p-\frac{1}{2}\right)}_{<0}<0 \qquad p\in[0;1]$$

$$p-\frac{1}{2}<0 \qquad\qquad p\in\left(0;\frac{1}{2}\right)$$

$$p<\frac{1}{2} \qquad\qquad p<\frac{1}{2}$$