2021-11-10

Bias - Variance Decomposition

Y target · $\quad (X, Y)$ από κοινό πληθυσμό

X predictors

Training set $T = \{(x_1, y_1), \ldots, (x_N, y_N)\}$

$T \Longrightarrow \underline{\underline{\hat{f}(x)}}$: prediction function

$\hat{y}_0 = \hat{f}(x_0; T)$ $\qquad$ x_0: νέα τιμή του X

"Σφάλμα πρόβλεψης"

$X = x_0 \Longrightarrow \hat{f}(x_0)$: πρόβλεψη

Σφάλμα $\left(\hat{f}(x_0) - \underbrace{\text{"true"}}_{?}\right)^2$

$\underline{Y|X} = f(X) + \varepsilon$ $\qquad$ ε : τυχαία απόκλιση

$E(Y|X=x_0) = f(x_0)$ $\qquad$ $E(\varepsilon)$, $Var(\varepsilon) = \sigma^2$

① Σφάλμα λόγω του ε

Εκτίμηση του $E(Y|X=x_0) = f(x_0) \leftarrow \underline{\underline{\Delta E}}$

$E(Y|X=x_0)$

Πρόβλεψη ενός μελλοντικού Y για $X = x_0$.

② $\hat{f}(x_0) \neq \overset{E(Y|X=x_0)}{f(x_0)}$

$\hat{f}(x_0) \neq E\hat{f}(x_0)$?

$E\hat{f}(x_0) \neq f(x_0)$

$f(x_0) = E(Y|X=x_0)$ άγνωστο.

$T_1 \Rightarrow \hat{f}_1(x_0)$

$T_2 \Rightarrow \hat{f}_2(x_0)$

$\vdots$

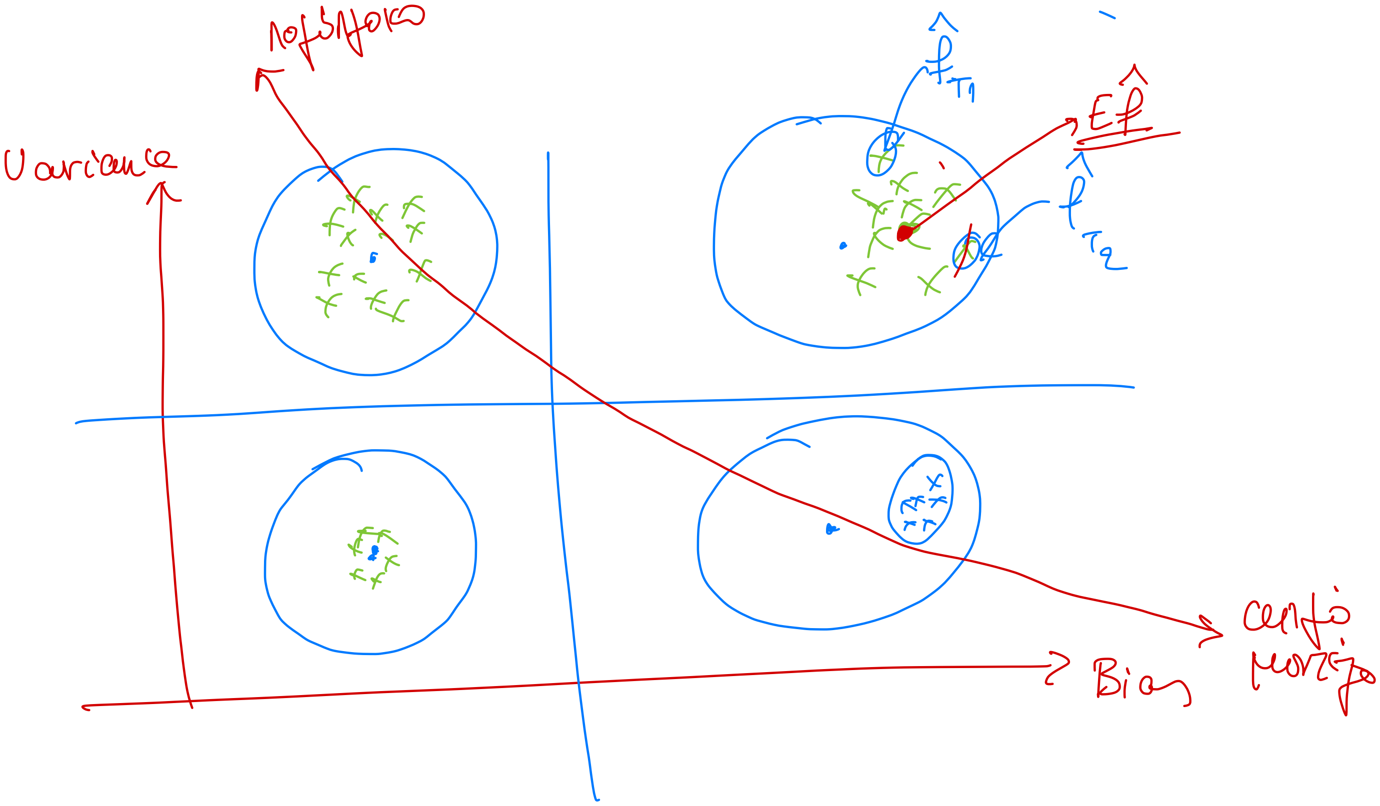
$E\hat{f}(x_0)$

μέση πρόβλεψη του $f(x_0)$ ως προς διαφορ. T

$f(x_0)$ — $E\hat{f}(x_0)$ — $\hat{f}(x_0)$

true. — μέση πρόβλεψη. — από ένα T.

bias — variance, διασπορά του $\hat{f}(x_0)$ ως προς T



Y εξαρτ. μετ. (target)

X ανεξ. μετ. (predictors–inputs)

T = training set $\begin{pmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_N & y_N \end{pmatrix} \Longrightarrow \underline{\underline{\hat{f}(x)}}$: prediction function

↳ π.χ. $\underline{\hat{f}(x)} = \underline{5x_1^2 + 2\log(x_1+x_2)}$

Συνάρτηση $L(Y, \underline{\hat{f}(X)})$

---

① <u>Test Error</u> (Generalization Error)

$$Err_T = E\left[L(Y, \hat{f}(X)) \mid T \right]$$

μέση τιμή ως προς την από κοινού (X, Y)

το training set θεωρείται <u>γνωστό και σταθερό</u>.

$\underline{Err_T}$: αναμενόμενο σφάλμα (απώλεια) σε <u>νέα δεδομένα</u> (X, Y) χρησιμοποιώντας για πρόβλεψη την $\hat{f}(\cdot)$ που έχει προέλθει από το συγκεκριμένο T.

② Expected Prediction Error (Μέσο Σφάλμα πρόβλεψης)

$$Err = E(Err_T)$$

μέση τιμή ως προς διαφορετικά T.

$$= E\left(L(Y, \hat{f}(X)) \right)$$

→ λαμβάνει υπόψη <u>όλη την αβεβαιότητα του συστήματος</u> (T, νέο (X, Y))

③ Training Error

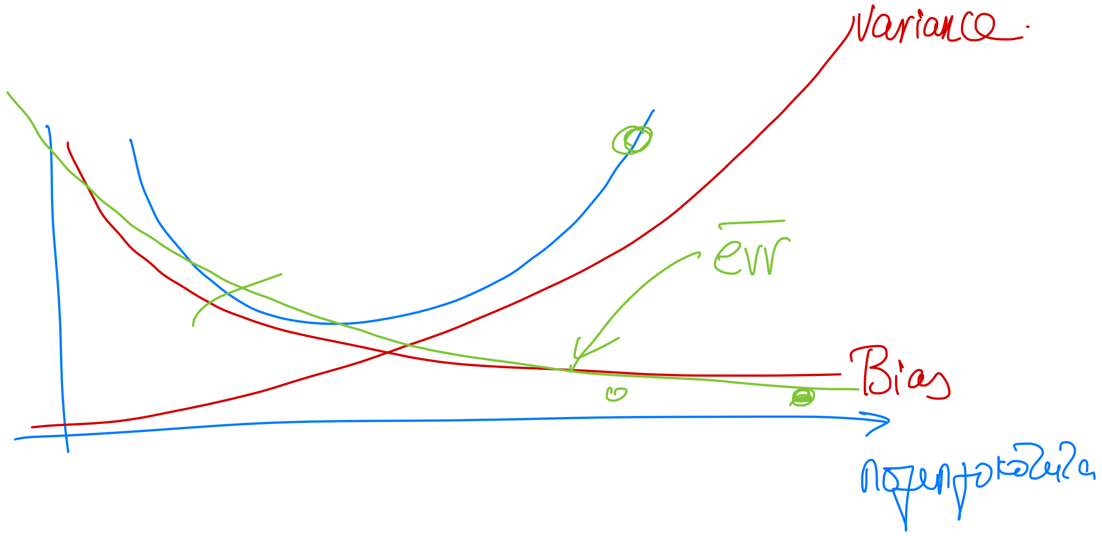
$$\overline{err} = \frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{f}(x_i))$$

παίρνω το συγκεκριμένο T.

$$T \; \{(x_i, y_i),\ i=1,\dots,N\} \Longrightarrow \hat{f}(\cdot) \Longrightarrow \left.\begin{matrix} L(y_1, \hat{f}(x_1)) \\ L(y_2, \hat{f}(x_2)) \\ \vdots \\ L(y_N, \hat{f}(x_N)) \end{matrix}\right\} \overline{err}$$

$\overline{err} \searrow 0$ με την πολυπλοκότητα.

$\left.\begin{matrix} Err_T \\ Err \end{matrix}\right\}$



Dataset

Training Set	Validation Set	Test Set
fitting various models (π.χ. παραμέτρων)	επιλογή μοντέλων.	εκτίμηση σφάλματος του καλύτερου μοντέλου

καταλήγουμε στο "καλύτερο" μοντέλο