7/3/2025

## Supervised Learning

$Y$ : εξαρτημένη μεταβλητή / response/outcome

Προβλέψεις για την $Y$ με βάση προηγούμενα δεδομένα

Πλαίσιο

$X = (X_1, \dots, X_p)$ ανεξάρτητες μεταβλητές/predictors /features

$Y$ = outcome

$H_X(x)$ : περιθώρια κατανομή του $X$

$H_{Y|X}(y|x) = P(Y \leq y \mid X = x)$ δεσμ. cdf $Y|X=x$.

$f(x) = E(Y|X=x) = \int_x y \, dH(y|x)$

Πρόβλημα: Δεδομένου $X=x \rightarrow$ Πρόβλεψη για $Y=?$

Έστω $p(x)$ = Πρόβλεψη του $Y$ για $X=x$.

Κριτήριο (συνάρτηση απώλειας) : "min" $L(Y, p(x))$ : ποινή αν πρόβλεψη : $p(x)$, πραγμ. τιμή : $Y$

Ιδιότητα $L \geq 0$ $\forall y, p(x)$

$L(y, y) = 0$ $\forall y$.

$L(y, c) > 0$ αν $c \neq y$

δεν έχει νόημα το πρόβλημα $\min_c L(y,c)$

$\min_c L(y,c) = 0$ για $c=y$

Θεωρούμε $\min_c E_{Y|X=x}\left(L(Y,c)\right)$
$(X=x)$

Ποια $L$ ?

① $Y$ : συνεχής (regression)

$$L(y,c) = (y-c)^2$$

$$\min_c E_{Y|X}\left[(Y-c)^2\right] =$$

$$=\min_c \left\{ E_{Y|X}\left(Y^2 - 2Yc + c^2\right)\right\} =$$

$$=\min_c \left\{ E(Y^2|X=x) - 2cE(Y|X=x) + c^2 \right\}$$

min για $\boxed{c = E(Y|X=x) = f(x)}$

$$\min EL = E_{Y|X=x}\left[\left(Y-f(x)\right)^2\right]$$

$$= E_{Y|X=x}\left[\left(Y - E(Y|X=x)\right)^2\right] =$$

$$= Var(Y|X=x) \leftarrow \text{irreducible error}$$

**Πρόβλημα** $H(y|x)$ : άγνωστη

$\Rightarrow$ $f(x)$ : άγνωστη συνάρτηση

χρειαζόμαστε δεδομένα

Δεδομένα [Σύνολο εκπαίδευσης (training set)]

τυχαίο δείγμα μεγέθους $n$:

| $i$ | $x_{1\cdot}$ | $x_{2\cdot}$ | $\cdots$ | $x_{p\cdot}$ | $y$ |
|---|---|---|---|---|---|
| 1 | $x_{11}$ | $x_{21}$ | $\cdots$ | $x_{p1}$ | $y_1$ |
| 2 | $\vdots$ | | | | |
| $\vdots$ | $\vdots$ | | | | |
| $n$ | $x_{1n}$ | $x_{2n}$ | $\cdots$ | $x_{pn}$ | $y_n$ |

$x_{\cdot 1}$ ... $x_{\cdot n}$

$\Rightarrow$ εκτίμηση $\hat{f}(x)$ συνάρτηση πρόβλεψης

Για νέες παρατηρήσεις με $X = x_0$ η πρόβλεψη θα είναι $\hat{y} = \hat{f}(x_0)$

Πώς υπολογίζεται η $\hat{f}(x)$ ?

Ανάλογα με τη μέθοδο/μοντέλο που χρησιμοποιούμε (πχ. regression, regularized regr, neural nets, random forests,...)

Ελαχιστοποίηση του Loss function αρκετά πιο περίπλοκο

$$\min_{\hat{f}} E_{?}\left[L(Y, \hat{f}(x))\right] = MPE$$

Mean Prediction Error

Ορίζω $\hat{f}$

min E(MPE)

training set $(x, y)$ $\Rightarrow \hat{f}(x)$

τυχαιότητα

νέες παρατηρήσεις

$x_{01}$
$x_{02}$
$\vdots$
$x_{0k}$

testing set (σύνολο δεδ. ελέγχου)

$\hat{y}_{01} = \hat{f}(x_{01})$
$\vdots$
$\hat{y}_{0k} = \hat{f}(x_{0k})$

παρατήρηση

$y_{01}$
$\vdots$
$y_{0k}$

σφάλματα πρόβλεψης

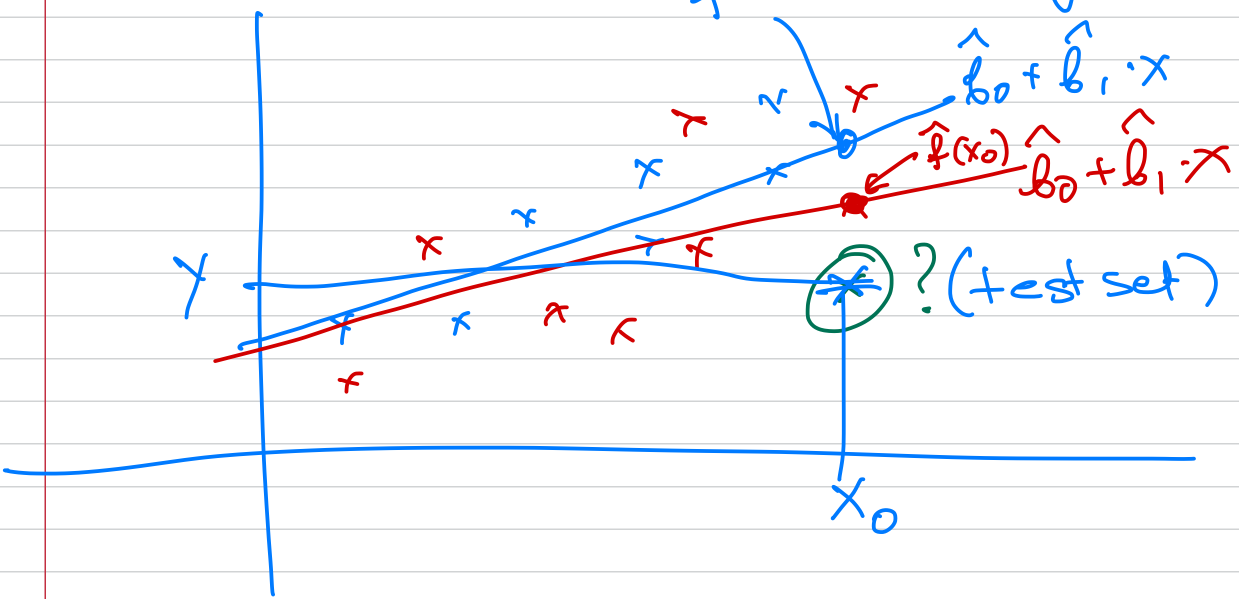
τυχαιότητα

π.χ. Regression για $p=1$ $\quad X = X_1$

$\hat f(x) = \hat b_0 + \hat b_1 \cdot x$ $\qquad$ $\hat b_0, \hat b_1$ : αντ. εξαρτ. τετρ.

τυχαίες μεταβλητές

εξαρτώνται από training set·




---

Μέσο σφάλμα πρόβλεψης

$$MPE(\hat f) = E_{X,Y}\left[\left(Y - \hat f(X)\right)^2\right]$$

Training set :

(a) $(x_1, \dots, x_n)$ : fixed εκ των προτέρων σχεδιασμένες (controlled studies) π.χ. $\underline{X} = (\overset{y_1}{25}, \overset{y_2}{25}, 25, 25, 25, 30, 30, \dots \dots \overset{y_n}{-})$

τυχαιότητα του training set: από $y_j, j=1,\dots,n$

(b) $(X_{\cdot 1}, \dots X_{\cdot n})$ τυχαία, ($y$ : τυχαία)

μελέτες παρατήρησης
(observational studies)

---

$$MPE = E_{Y,X}\left(\left(Y - \hat{f}(X)\right)^2\right)$$

$\boxed{X, Y : \text{test set}}$ !!

training set $\Longrightarrow$ $\boxed{\hat{f} : \text{τυχαία μεταβλητή}}$

$\hat{f}(X)$ → $X$ : τυχαίο (test set)
→ $\hat{f}$ : τυχαίο (προέρχεται από το training set)

# Παράδειγμα

Έστω δείγμα (training) $n=20$

$X = (x_1, \dots, x_n) \rightarrow$ fixed $(\in (0,10))$

$Y = (y_1, \dots, y_n)$

$$Y = \underbrace{2 + 3.6x + 1.5x^2 - 0.15x^3}_{f(x)} + \varepsilon$$

$\varepsilon \sim N(0, \sigma^2)$, $\sigma = 5$

Θεωρούνται άγνωστα από τη μέθοδο πρόβλεψης

Μέθοδος πρόβλεψης (Πολυωνυμική Παλινδρόμηση ελαχίστων τετραγώνων)

$$\hat{f}(x) = \hat{b}_0 + \hat{b}_1 x + \hat{b}_2 x^2 + \dots + \hat{b}_k x^k \quad (k: \text{τάξη μοντέλου})$$

$\hat{b}_0, \dots, \hat{b}_k$ : εκτιμ. ελαχ. τετραγώνων

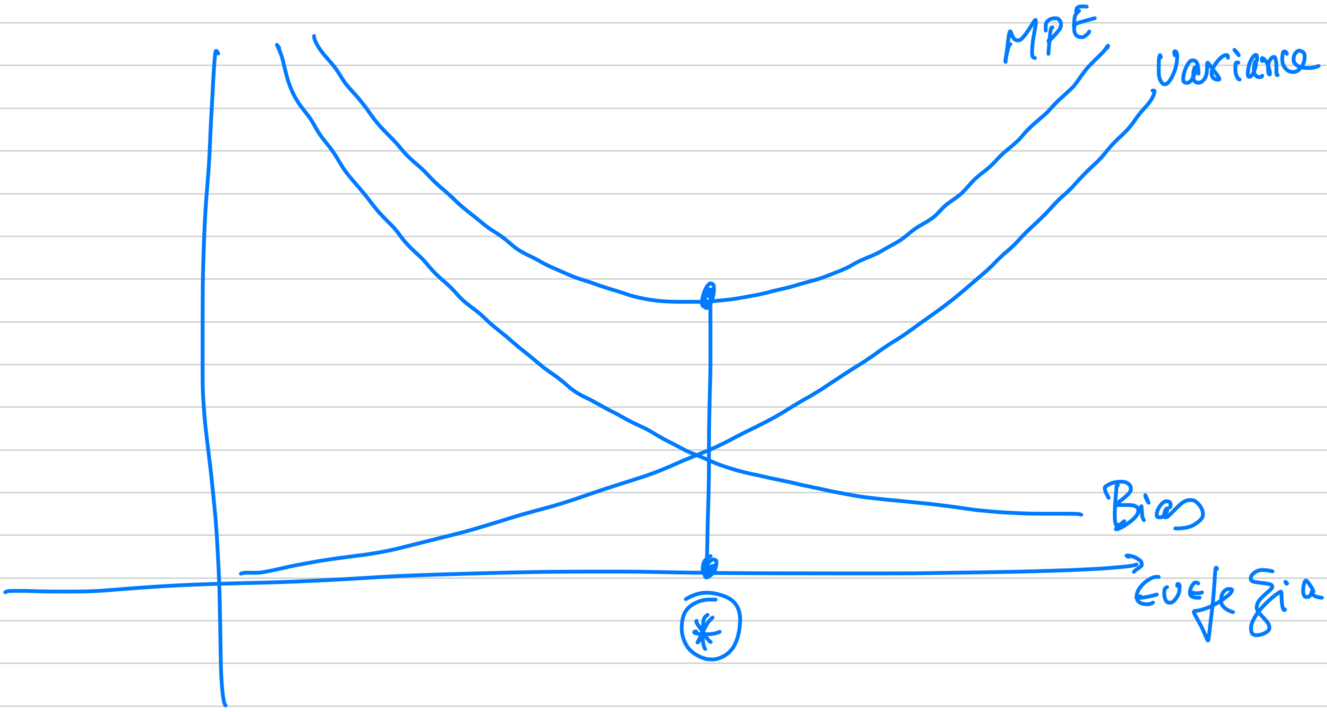
test : $x_0 = 6.2$

κ : ευελιξία του μοντέλου (flexibility)
(αρ. παραμέτρων που εκτιμώνται $(b_0, b_1, \dots, b_\kappa)$)

Σε <u>παραμετρικά μοντέλα</u>

π.χ. $\hat{f}(x) = b_0 + b_1 x$ ← γραμμικό μοντέλο (b)

$\quad = b_0 + b_1 \sin(b_2 x)$ μη γραμμικό (b)




MPE = . . . . .
(bias, variance)

$$MPE = E_{X,Y}\left(Y - \hat{f}(X)\right)^2$$

X, Y (test set)
(νέες παρατηρήσεις)

$\hat{f}$ : από training set

π.χ. έστω $X = x_0$ = σταθερό

$$Y \sim H(y \mid x_0) \quad = \quad E(Y) = f(x_0)$$

$\hat{f}$: από training set

$x_1, \dots, x_n$ : σταθερά

$y_1, \dots, y_n$ τυχαία $\quad y_i \sim H(\cdot \mid x_i)$

$\downarrow$

$\hat{f}(x_0)$: τυχαία μεταβλητή

$\hat{f}(x_0)$ <u>ανεξάρτητη</u> $\quad Y \sim H(Y \mid X_0)$

$\downarrow$ training set $\qquad \downarrow$ test

$$E\left(Y - \hat{f}(x_0)\right)^2 =$$

$$= E_{tr,Y}\left( \underline{Y - f(x_0)} + \underline{f(x_0)} - \underline{E(\hat{f}(x_0))} + \underline{E(\hat{f}(x_0))} - \underline{\hat{f}(x_0)} \right)^2$$

$f(x_0) = E(Y \mid X = x_0) \leftarrow H(y \mid x_0)$ (άγνωστη πραγματική μέση τιμή του $Y \mid X = x_0$)

$\underline{E(\hat{f}(x_0))}$ = μέση τιμή της πρόβλεψης για $X = x_0$

(από training set)

$$MPE = E\left( (Y - f(x_0)) - \left(\hat{f}(x_0) - E(\hat{f}(x_0))\right) + \left(f(x_0) - E(\hat{f}(x_0))\right) \right)^2$$

$$= E\Big[ (Y_0 - f(x_0))^2 + \left(\hat{f}(x_0) - E(\hat{f}(x_0))\right)^2 + \left(f(x_0) - E(\hat{f}(x_0))\right)^2$$

$$- 2(Y - f(x_0)) \cdot \left(\hat{f}(x_0) - E(\hat{f}(x_0))\right)$$

$$+ 2(Y - f(x_0))\left(f(x_0) - \underline{E(\hat{f}(x_0))}\right)$$

$$-2\left(\hat{f}(x_0)-E(\hat{f}(x_0))\right)\cdot\left(f(x_0)-E(\hat{f}(x_0))\right)\Big]$$

Παρατήρηση

$$E\left[(Y-f(x_0))(\hat{f}(x_0)-E(\hat{f}(x_0))\right]$$

$$=\underbrace{E(Y-f(x_0))}_{=0}\cdot E\Big[\underbrace{\hat{f}(x_0)-E(\hat{f}(x_0))}_{=0}\Big]=0$$

$$E\left((Y-f(x_0))(f(x_0)-E(\hat{f}(x_0))\right)$$

$$=\left[f(x_0)-E(\hat{f}(x_0))\right]\cdot \underbrace{E\left(Y-f(x_0)\right)}_{=0}=0$$

όμοια $E\Big[(\hat{f}(x_0)-E(\hat{f}(x_0))\cdot\underbrace{(f(x_0)-E(\hat{f}(x_0))}_{\text{σταθ.}}\Big]$

$=0$

$Y\sim H(y|x_0)$

$E(Y)=f(x_0)$

$$\Rightarrow MPE = E\left((Y-f(x_0))^2\right)$$

$$+\ E\left[(\hat{f}(x_0)-E(\hat{f}(x_0))^2\right]$$

$$+\ \left(E(\hat{f}(x_0))-f(x_0)\right)^2$$

$$= Var(Y \mid X = x_0) \quad \leftarrow \text{(irreducible error)}$$

$$+ \; Var(\hat{f}(x_0))$$

$$+ \; Bias^2$$

!!

$$\boxed{MPE = Bias^2 + Variance + \underbrace{\text{Irred. Error}}_{\text{ανεξ. } \hat{f}}}$$