Supervised learning

customer type	# of new recipients	email length	country	email type
gold	0	2	germany	—
silver	1	4	germany	—
Bronze	5	2	Nigeria	Spam
Bronze	2	4	russia	Spam
Bronze	3	4	USA	—
gold	0	1	USA	—

סוגי דאטה:

אורדינליים

נומריים

נומינליים

כל שורה נקראת instance

כל עמודה נקראת feature

התיוגים נקראים labels

לכל instance נקרא X ולכל X יהיה:
קיים label מסומן ב Y

נגדיר פונקציית "מחיר"

$$L = (h(x) - y)^2$$

$h(x)$ - ההיפותזה

L - מחיר

$$l(x, y, h) = (h(x) - y)^2$$

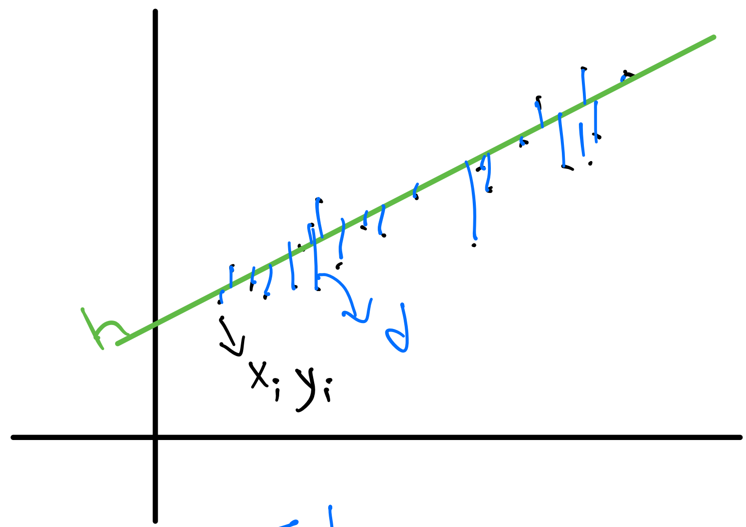
$$L_S = \frac{1}{m}\sum_{i}^{m} l(x_i, y_i, h) = \frac{1}{m}\sum_{i}^{m} (h(x_i) - y_i)^2$$

$$L_D = \mathop{E}_{x \sim D}[l(x_i, y_i, h)]$$

$$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_N x_N$$

יש כל מיני דרכים להגדיר Loss, כשעושים רגרסיה לינארית מחפשים למצוא מה יביא למינימום את מרחק הפיזורים מהמודל

למשל:



$S_d = \sum d \to \min \to (S_d)'$

לכן

$$L_S = \frac{1}{2m}\sum_i^m (h(x)-y)^2$$

2 רק. לבטל
את החזקה בגזירה

m דגימות

$x, y, h \leftarrow$ עבור כל דגימה

$$X = \begin{bmatrix} — x_1^T — \\ — x_2^T — \\ \vdots \\ — x_m^T — \end{bmatrix}$$

n פיצ'רים.

m דגימות (instances)

לכן, כיוון שהמרחק מהמודל (היפותזה) יכול להיות חיובי/שלילי $(h-y)$, מתייחסים לריבוע המרחק

$(h-y)^2$ כאשר h היא פונקציה של הפיצ'רים.

הנחה לרגרסיה:

$$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

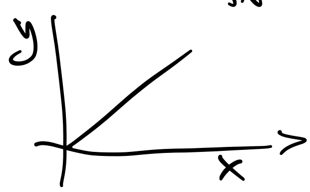
θ_0: bias

$\theta_1 x_1$: פיצ'ר 1

$\theta_2 x_2$: פיצ'ר 2

$\theta_n x_n$: פיצ'ר n

במקרה של גרף דו מימדי



משתמשים במודל

$$h = \theta_0 + \theta_1 x$$

כיוון שבפונקציה יש פיצ'ר אחד (מערכת מפיצ'ר)

$$L = \frac{1}{2m}\sum^{m}(h-y)^2$$

$$h = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n \Rightarrow \begin{matrix} \theta_1^T x_1 \\ \theta_2^T x_2 \\ \vdots \end{matrix} = \underline{\underline{X}} \cdot \underline{\theta}$$

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ \vdots & & \ddots & & \\ x_{m1} & & \dots & & x_{mn} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

$-$

$mx1$ ו- $nx1$, mxn

כאן וקטורי העמודה.

$$L_S = \frac{1}{2m}\|\underline{\underline{X}}\,\underline{\theta} - \underline{y}\|^2 = \frac{1}{2m}\underbrace{(\underline{\underline{X}}\,\underline{\theta} - \underline{y})^t}_{1xm}\underbrace{(\underline{\underline{X}}\,\underline{\theta} - \underline{y})}_{mx1}$$

נגזור לפי וקטור $\underline{\theta}$ "לפי חוקי נגזרות רגילים". - לא נכון, יש כמה חוקים נוספים...

$$\frac{\partial}{\partial\underline{\theta}}\left(\frac{1}{2m}(\underline{\underline{X}}\,\underline{\theta}-\underline{y})^t(\underline{\underline{X}}\,\underline{\theta}-\underline{y})\right) =$$

$$\frac{\partial}{\partial\underline{\theta}}\left(\frac{1}{2m}\Big[\underbrace{(\underline{\theta}^t\underline{\underline{X}}^t - \underline{y}^t)(\underline{\underline{X}}\,\underline{\theta} - \underline{y})}\Big]\right) = 0$$

$$\Rightarrow \underset{1xn\ \ nxm\ \ mxn\ \ nx1}{\underline{\theta}^t\underline{\underline{X}}^t\underline{\underline{X}}\,\underline{\theta}} - \underbrace{\overset{1xn\ \ nxm\ \ mx1}{\underline{\theta}^t\underline{\underline{X}}^t\underline{y}}}_{1x1} - \underbrace{\overset{1xm\ \ mxn\ \ nx1}{\underline{y}^t\underline{\underline{X}}\,\underline{\theta}}}_{1x1} + \underset{1xm\ \ mx1}{\underline{y}^t\underline{y}} = 0$$

$$\to \frac{\partial}{\partial\underline{\theta}} = \quad 1x1$$

$$2\underline{\underline{X}}^t\underline{\underline{X}}\,\underline{\theta} - \underline{X}^t\underline{y} - \underline{X}^t\underline{y} + 0 = 0$$

$$\cancel{2}\underline{\underline{X}}^t\underline{\underline{X}}\,\underline{\theta} = \cancel{2}\underline{X}^t\underline{y}$$

$$\underline{\theta} = \underbrace{(\underline{\underline{X}}^t \underline{\underline{X}})}^{-1} \underline{x}^t \underline{y}$$

משתמשים ב pinv
כי לא תמיד יש inv

https://youtu.be/FCWrduAxf-Q

Common identities:

$$\frac{\partial(\boldsymbol{u}(\mathbf{x}) + \boldsymbol{v}(\mathbf{x}))}{\partial \mathbf{x}} = \frac{\partial \boldsymbol{u}(\mathbf{x})}{\partial \mathbf{x}} + \frac{\partial \boldsymbol{v}(\mathbf{x})}{\partial \mathbf{x}}$$

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\top$$

$$\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a}$$

$$\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^\top)\mathbf{x}$$

$$\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} \text{ if } \mathbf{A} \text{ is symmetric}$$

$$\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = |\mathbf{X}|(\mathbf{X}^{-1})^\top$$

$$\frac{\partial \ln |\mathbf{X}|}{\partial \mathbf{X}} = (\mathbf{X}^{-1})^\top$$