Notation

course grade (%) = Y

$$= 0.5 X_m + 0{,}15 X_Q + 0{,}15 X_{HW} +$$

$$+ 0{,}2 X_F$$

truth

$$\boxed{Y = f(X)}$$

Goal of Machine Learning

Find a good estimate of f, i.e. a good $\hat{f}$

Y is ctns then finding $\hat{f}(X)$ is called "regression"

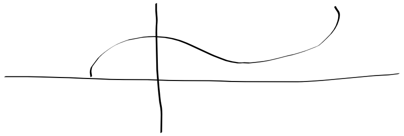
Y is discrete then it's "classification"

non-parameteric

few assumptions of the form of f

differentiable

Ex: $\widehat{\text{coursegrade}}$ = continuous function



What is a "good" $\hat{f}$?

$$|Y-\hat{Y}| \text{ is small}$$

dataset

$$\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = (\vec{y}-\hat{\vec{y}})^T(\vec{y}-\hat{\vec{y}})$$

$$\underset{1\times n}{\vec{y}^T = [y_1\, y_2 \cdots y_n]} \qquad \underset{n\times 1}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}} = \vec{y}$$

Matrix Form for dataset

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \vec{X} = \begin{bmatrix} X_{Q1} & X_{HW1} & X_{m1} & X_{F1} \\ X_{Q2} & X_{HW2} & X_{m2} & X_{F2} \\ \vdots & & & \\ X_{Qn} & X_{HWn} & X_{mn} & X_{Fn} \end{bmatrix}$$

n X 1 vector

n × ④ = pices of info per student

How to find $\hat{f}$?

training dataset $S = \{(y_1, \vec{x}_1),$

$(y_2, \vec{x}_2),$

$\vdots$

$(y_n, x_n)\}$

pick a form / method for fitting $\hat{f}$

$$\hat{y} = \hat{f}(x) = mx + b$$

minimize $\sum_{i=1}^{n} |y_i - \hat{y}_i|$

find m, b to

$$= \sum_{i=1} |y_i - (mx_i + b)|$$

Reality = true

$$\vec{X}\left\{\begin{array}{l} X_{HW} = \% \text{ earned on HW} \\ X_{Q} = \% \quad \text{'' \quad '' \quad Quizzes} \\ X_{m} = \text{'' \quad '' \quad '' \quad midterm} \\ X_{F} = \text{'' \quad '' \quad '' \quad final} \end{array}\right.$$

Y = course grade earned (%)

Prediction model

predicted course grade

$$= 62.5 + 0.21(\text{Quiz1})$$

$$\hat{Y} = 62.5 + 0.21(X_{Q1})$$

$$\hat{Y} = \hat{f}(X)$$

$\hat{f}$ is an estimate of f

Y = random variable

output

Y = random variable	X = random variable
output	input
response	predictors
	features
dependent	independent

parametric

specific form for f;

$$\hat{f}(x) = ax+b$$

or

$$ax^2+bx+c$$

Find $\hat{f}$ from

both Y, X
examples
||
supervised
learning

only X
||
unsupervised
learning

"datapoints"

= observations of X, Y

Notation: lower case

$$n \text{ examples} \left\{ \begin{array}{l} y_1, \vec{x}_1 = (x_{HW}, x_m, x_f, x_a)^T \\ y_2, \vec{x}_2 \\ \vdots \\ y_n, \vec{x}_n \end{array} \right.$$