Lesson B01 (a)
Regression

1. Definition of Regression.
2. Supervised Learning vs. unsupervised Learning vs. Reinforcement Learning.
3. Well-fitting vs. Overfitting vs. underfitting.
4. Linear Regression
5. Logistic Regression.

Regression Analysis.

① A statistical process for estimating the relation (Function) between a depedent variable (or label / outcome variable / output) and one or multiple independent variable (input / feature / predictor)

$\overbrace{x_1 \quad x_2 \quad x_3 \quad \cdots \quad x_n}^{\vec{x}}$	$\vec{y} = (y_1, y_2, \cdots y_m)$
input	output

② Given dataset $(\vec{X}, \vec{y})$
Disclose a function $f: R^n \to R^m$.
to fit the dataset.

③ Purposes: $\begin{cases} \text{prediction.} \\ \text{to infer relationship.} \end{cases}$

Supervised Learning vs.
Unsupervised Learning vs.
Reinforcement Learning.

Supervised Learning (Using Ground Truth)

① Given a set of N training example

$\{(\vec{x}_1, \vec{y}_1), (\vec{x}_2, \vec{y}_2), \ldots (\vec{x}_N, \vec{y}_N)\}$.

seek a function $H(x) : X \to Y$.

(hypothesis)

input space, output space

$\vec{x} \in X$. $\vec{y} \in Y$.

② such that specific Objective function is optimized!

classification + Regression

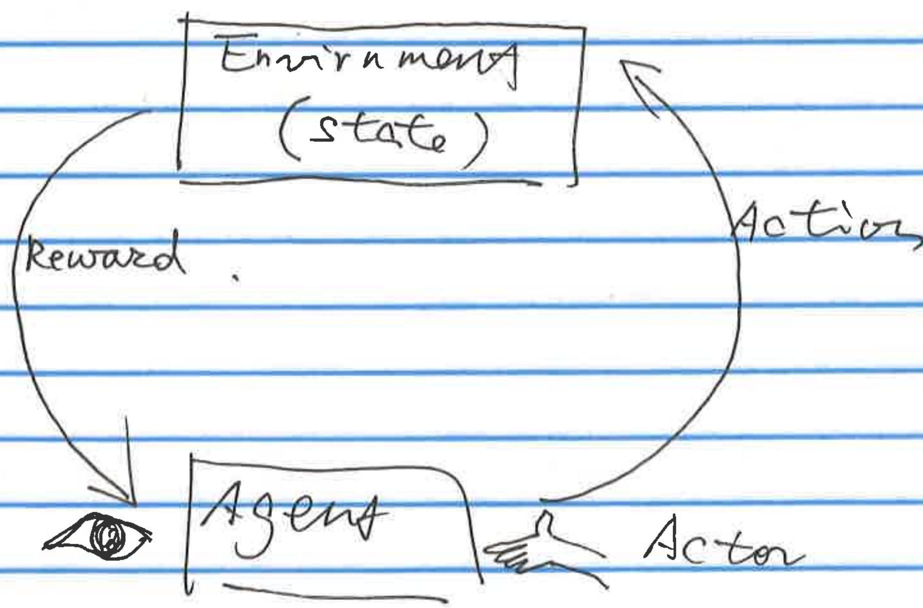
Unsupervised Learning.

→ No labels are provided.

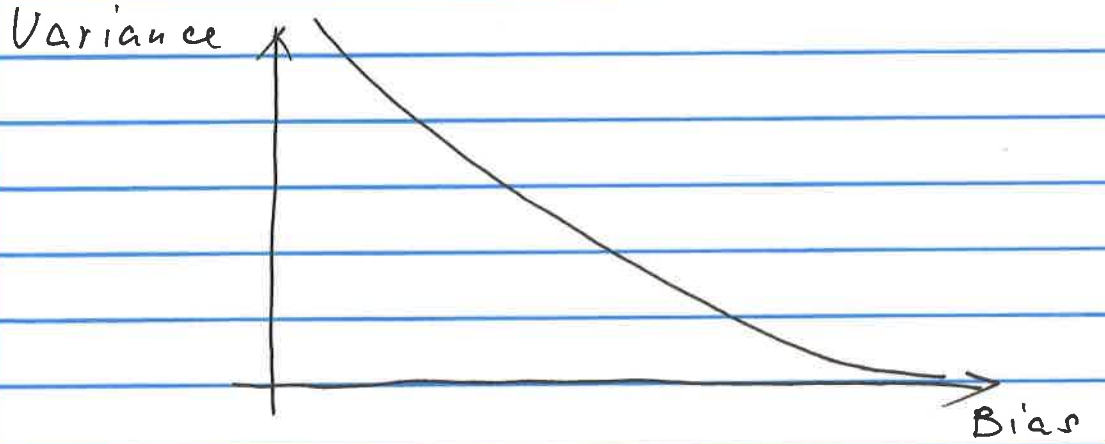
→ Representative Tasks

- Clustering: KNN, K-means
- Dimensionality Reduction: PCA.
- Exploratory analysis: Autoencoder.

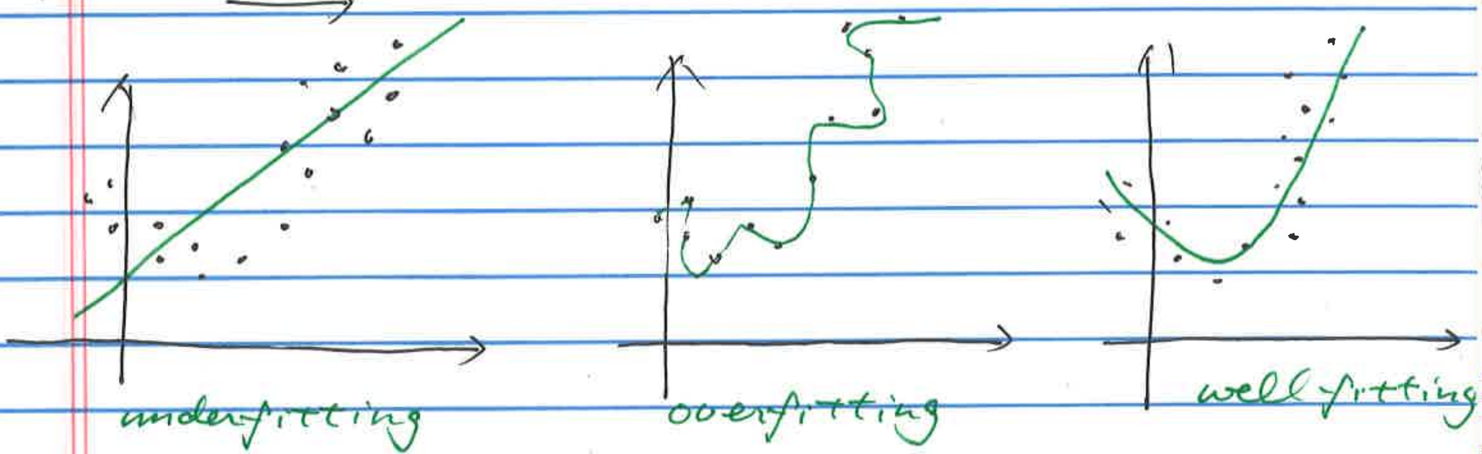
Reinforcement Learning.



Bias – variance Tradeoff.

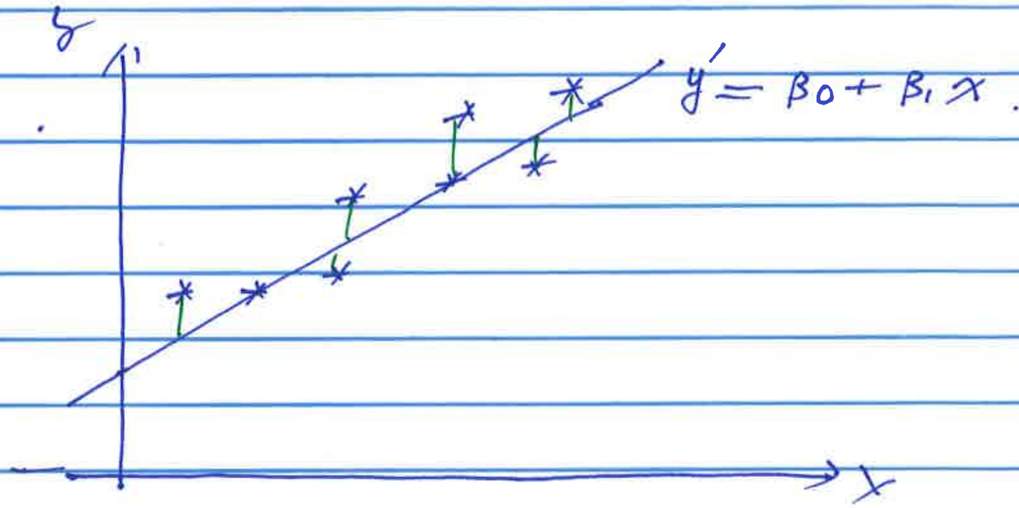


Lower bias in parameter estimation



* Bias: an Error from erroneous Assumption of Learning Alg (underfitting)
* variance: an Error from small fluctions in the training data (overfitting)

Derivation of Linear Regression



Given (x_1, y_1) (x_2, y_2) $\cdots$ (x_N, y_N).
Find the best-fit straight line:

$$y'(x) = \beta_0 + \beta_1 x.$$

residual $r_i = y'(x_i) - y_i$

Sum of the square of residual

$$S(\beta_0, \beta_1) = \sum_{i=1}^{N} (\beta_0 + \beta_1 x_i - y_i)^2.$$

To minimize $S(\beta_0, \beta_1) \Rightarrow \frac{\partial S}{\partial \beta_0} = 0, \frac{\partial S}{\partial \beta_1} = 1$

(Stationary Point)

because $\nabla^2 S$ is positive definite
(Hessian)

$$\begin{cases} \dfrac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = \displaystyle\sum_{i=1}^{N} 2(\beta_0 + \beta_1 x_i - y_i) = 0. & (1) \\[2ex] \dfrac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = \displaystyle\sum_{i=1}^{N} 2(\beta_0 + \beta_1 x_i - y_i) \cdot x_i = 0. & (2) \end{cases}$$

$N \cdot \beta_0$.

$$\begin{matrix} (1) \Longrightarrow \\ (2) \Longrightarrow \end{matrix} \begin{cases} \displaystyle\sum_{i=1}^{N} \beta_0 + \sum_{i=1}^{N} \beta_1 x_i = \sum_{i=1}^{N} y_i & (3) \\[2ex] \displaystyle\sum_{i=1}^{N} \beta_0 x_i + \sum_{i=1}^{N} \beta_1 x_i^2 = \sum_{i=1}^{N} x_i y_i & (4) \end{cases}$$

$$\Longrightarrow \begin{cases} N\beta_0 + \beta_1 \displaystyle\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i \\[2ex] \beta_0 \cdot \displaystyle\sum_{i=1}^{N} x_i + \beta_1 \sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} x_i y_i \end{cases}$$

$$\Longrightarrow \begin{bmatrix} N & \displaystyle\sum_{i=1}^{N} x_i \\[2ex] \displaystyle\sum_{i=1}^{N} x_i & \displaystyle\sum_{i=1}^{N} x_i y_i \end{bmatrix} \begin{bmatrix} \beta_0 \\[2ex] \beta_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\[2ex] \sum x_i y_i \end{bmatrix}.$$

$$\Rightarrow \begin{cases} \beta_0 = \dfrac{\sum_{i=1}^{N} y_i}{N} - \beta_1 \dfrac{\sum x_i}{N} \\ \quad = \bar{y} - \beta_1 \bar{x} \\ \beta_1 = \dfrac{N \cdot \sum_{i}^{N} (x_i y_i) - \sum_{i=1}^{N} x_i \cdot \sum_{i=1}^{N} y_i}{N \cdot \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2} \end{cases}$$

$\boxed{\text{end}}$

Multiple Linear Regression.

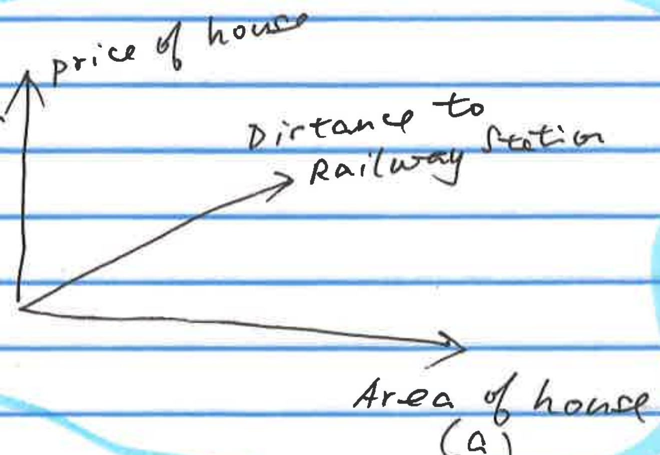
— More than one explanatory variables are involved.

y-intercept slope

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 \cdot x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon .$$

dependent variable

Exercise:



Linear Regression in Vector Format

x_1	x_2	$\dots$	x_p	y	y'
5	2	$\dots$	3	7	$\beta_0 + \beta_1 x_1 \dots \beta_p x_p$
1	2.0		4.0	8	
.	-	-	.		
$\vec{x}_1$	$\vec{x}_2$		$\vec{x}_3$	$\vec{y}$	$\vec{y}'$

$$\overrightarrow{\text{residual}} = \vec{y}' - \vec{y}$$

$$= \beta_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \vec{x}_1 + \beta_2 \vec{x}_2 + \dots + \vec{x}_p \beta_p - \vec{y}$$

the $\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$ is formulated by

$$\underset{\begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix}}{\operatorname{argmin}} \left\| \overrightarrow{\text{residual}} \right\|^2$$

square of norm