Korea Aerospace University

2AI4203: Machine Learning

Lec02: Learning from Data & Linear Regression

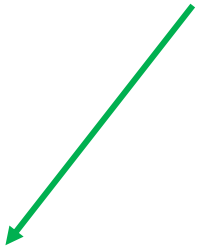
Jiyoon Shin

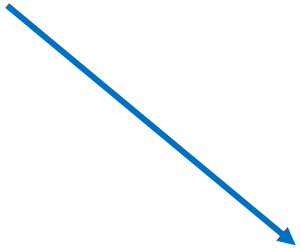
KAU College of AI Convergence

Sep. 09, 2026

Machine Learning

Machine Learning: **Learning** patterns from **Data**





Part 2. Learning
- Linear Regression
- Loss Function
- Gradient Descent

Part 1. Data
- Data, Dataset
- Supervised/Unsupervised Learning
- Classification & Regression

Korea Aerospace University

Part 1. Data

Data

- A data point can be represented as an input-output pair:

$$(X, Y)$$

- X: Input / Feature / Predictor / Attribute
- Y: Output / Label / Target

- Different sources may use different terms for X and Y


Cat


Cat


Dog


Dog

Example Input-Output Pairs

- Predict a student's GPA from exam scores
  - Features = scores on 5 exams
  - Target = GPA
  - Example input-output pair:

$$\big((85, 96, 72, 66, 88), 1.33333\big)$$



Example Input-Output Pairs

- Predict whether a sentence is a lie
  - Input = A statement made by a person
  - Output = Whether the sentence was truthful or a lie
  - Example input-output pair:

("I am not a crook.", "lie")

Statement X

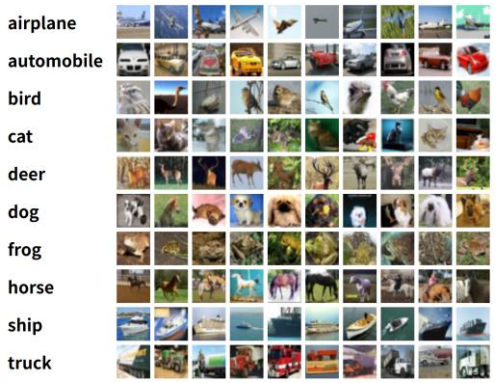
Truth/Lie Y

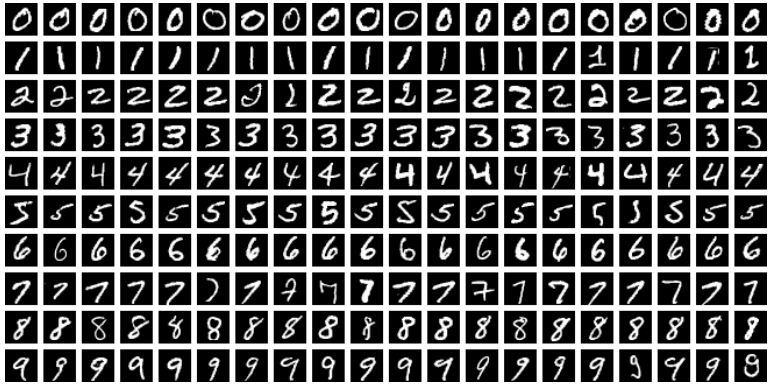
Dataset

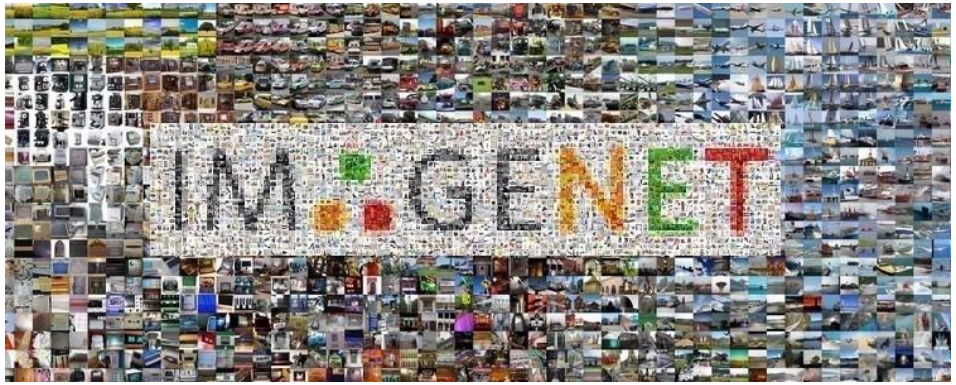
- A dataset is a collection of n data points:

$$D = \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$$

- X_i: input of the i^{th} example
- Y_i: output of the i^{th} example
- n: number of examples

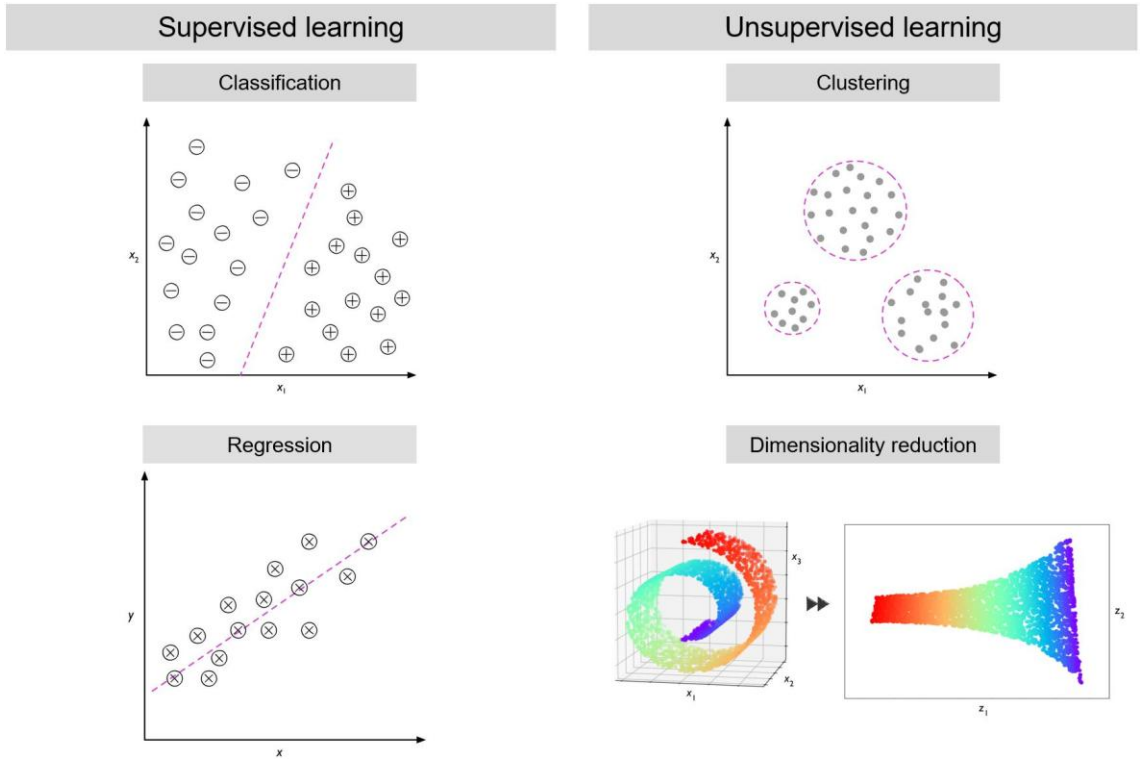






Supervised/Unsupervised Learning

- **Do we always have labels?**
- Supervised Learning
  - Learn with labels
  - Data: (X, Y)
  - Common tasks: classification, regression
- Unsupervised Learning
  - Learn without labels
  - Data: X
  - Common tasks: clustering, dimensionality reduction
- Other learning paradigms will be introduced later

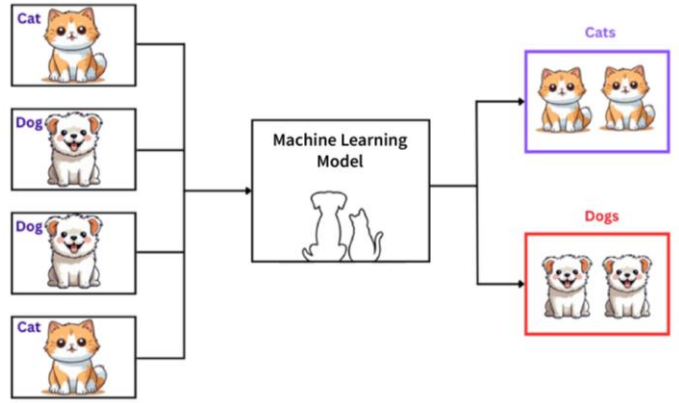


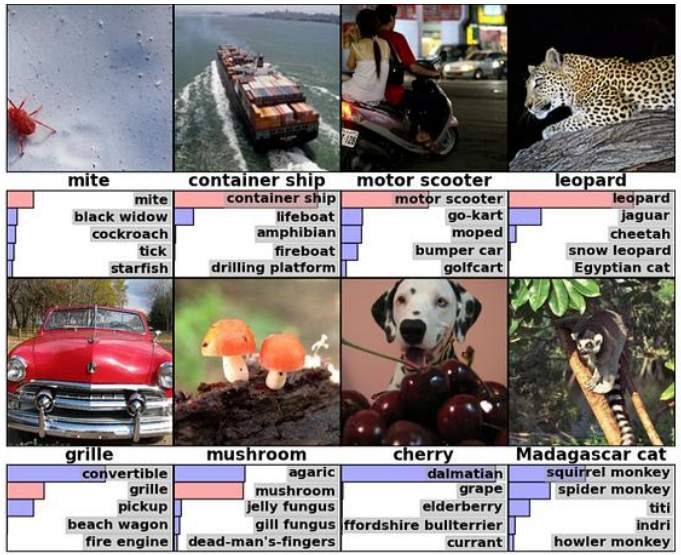
Classification

- Predict a category (class)

- Binary classification: 2 classes
  - Cat / Dog
  - Spam / Not spam
- Multiclass classification: 3 or more classes
  - Cat / Dog / Bird / …
  - 1 / 2 / 3 / …

- $Y \in \{\text{class}_1, \text{class}_2, \ldots\}$





Regression

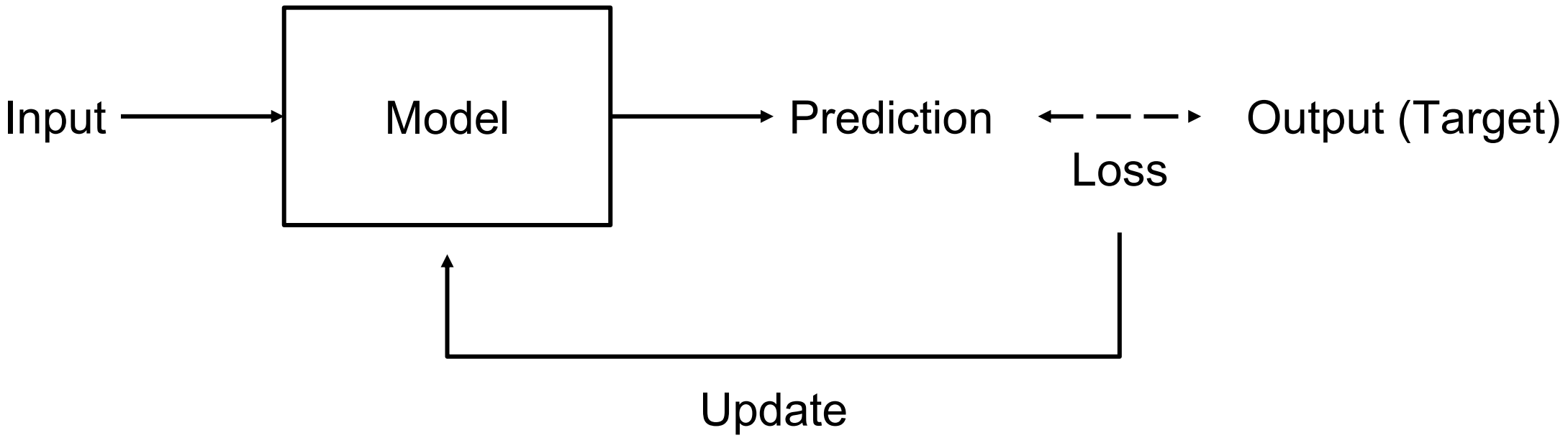
- Predict a continuous value

- Example
  - House features → Price: $420K
  - Weather data → 24.5
  - Exam scores → GPA: 3.72

- $Y \in \mathbb{R}$

Variables	(1) *CR*	(2) *IJE*	(3) *IJE-CR*	(4) *Macro*	(5) *Firm*	(6) *Firm*	(7) *Firm All*	(8) *Fama McBeth*	(9) *Firm All*
CR	−0.021***		−0.016***	−0.018***	−0.014	−0.018**	−0.018**	−0.019***	−0.020***
	(0.007)		(0.003)	(0.004)	(0.009)	(0.007)	(0.007)	(0.002)	(0.001)
IJE		0.109*	0.095**	0.100***	0.119*	0.131*	0.135**	0.132***	0.126***
		(0.059)	(0.042)	(0.030)	(0.066)	(0.066)	(0.064)	(0.005)	(0.006)
PROF					−0.002	−0.019	−0.036	−0.038***	−0.005
					(0.016)	(0.026)	(0.036)	(0.007)	(0.007)
SZ					0.006*	0.007*	0.007*	0.008***	0.006***
					(0.003)	(0.004)	(0.004)	(0.000)	(0.000)
COLAT					0.118***	0.125***	0.115***	0.115***	0.126***
					(0.018)	(0.017)	(0.016)	(0.006)	(0.005)
RD					−0.070	−0.069	−0.056	−0.067***	−0.057***
					(0.073)	(0.080)	(0.096)	(0.011)	(0.022)
LIQ					−0.005***	−0.005***	−0.007***	−0.007***	−0.009***
					(0.001)	(0.001)	(0.002)	(0.000)	(0.000)
TAX					−0.038***	−0.034***	−0.034***	−0.035***	−0.045***
					(0.012)	(0.012)	(0.012)	(0.002)	(0.004)
CAPEX					0.067	0.059	0.113	0.114***	0.042***
					(0.066)	(0.059)	(0.078)	(0.020)	(0.016)
DCPSF				−0.010		−0.020	−0.021	−0.023***	−0.019***
				(0.015)		(0.017)	(0.016)	(0.005)	(0.002)
MCAP				0.008		0.017*	0.018*	0.022***	0.022***
				(0.007)		(0.010)	(0.010)	(0.004)	(0.002)
VOL							−0.092***	−0.096***	−0.127***
							(0.028)	(0.009)	(0.015)
ASLF					−0.009				
					(0.047)				
Constant	0.169***	0.112***	0.160***	0.159*	0.093**	0.095	0.118	0.103***	0.109**
	(0.026)	(0.020)	(0.020)	(0.082)	(0.041)	(0.073)	(0.076)	(0.021)	(0.043)
Observations	303,270	189,570	189,290	181,109	131,176	126,004	100,703	100,703	100,703
R-squared	0.129	0.127	0.139	0.138	0.218	0.221	0.227	0.239	0.268
Industry dummy	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year dummy	YES	YES	YES	YES	YES	YES	YES	YES	YES

Machine Learning Workflow



Machine Learning (revisited)

Machine Learning: **Learning** patterns from **Data**



Part 2. Learning
- Linear Regression (Model)
- Loss Function (Loss)
- Gradient Descent (Update)

Part 1. Data
- Data, Dataset
- Supervised/Unsupervised Learning
- Classification & Regression

Korea Aerospace University

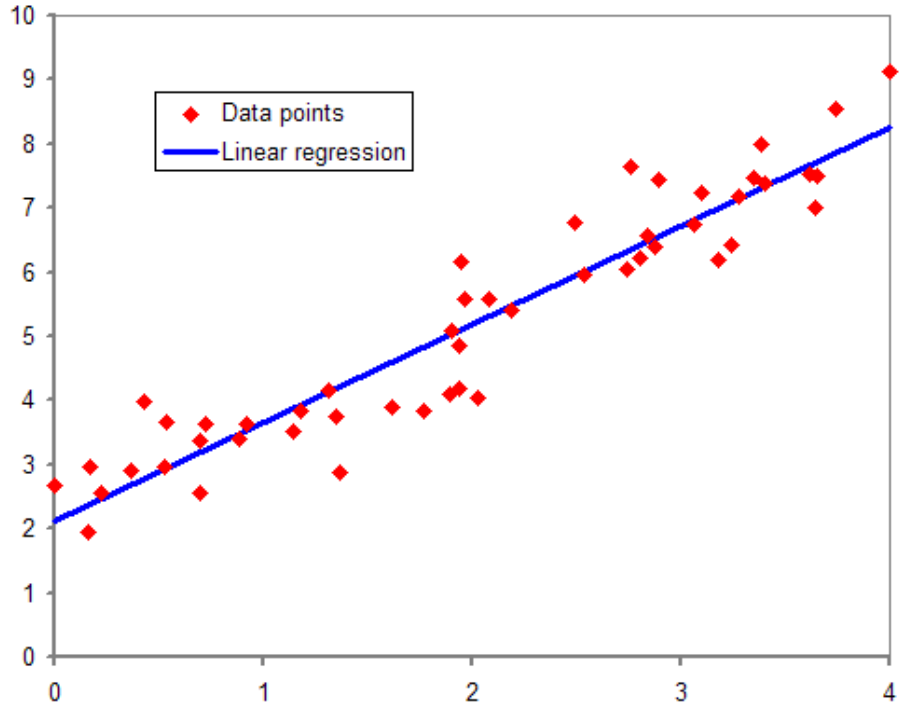
Part 2. Learning

Linear Regression

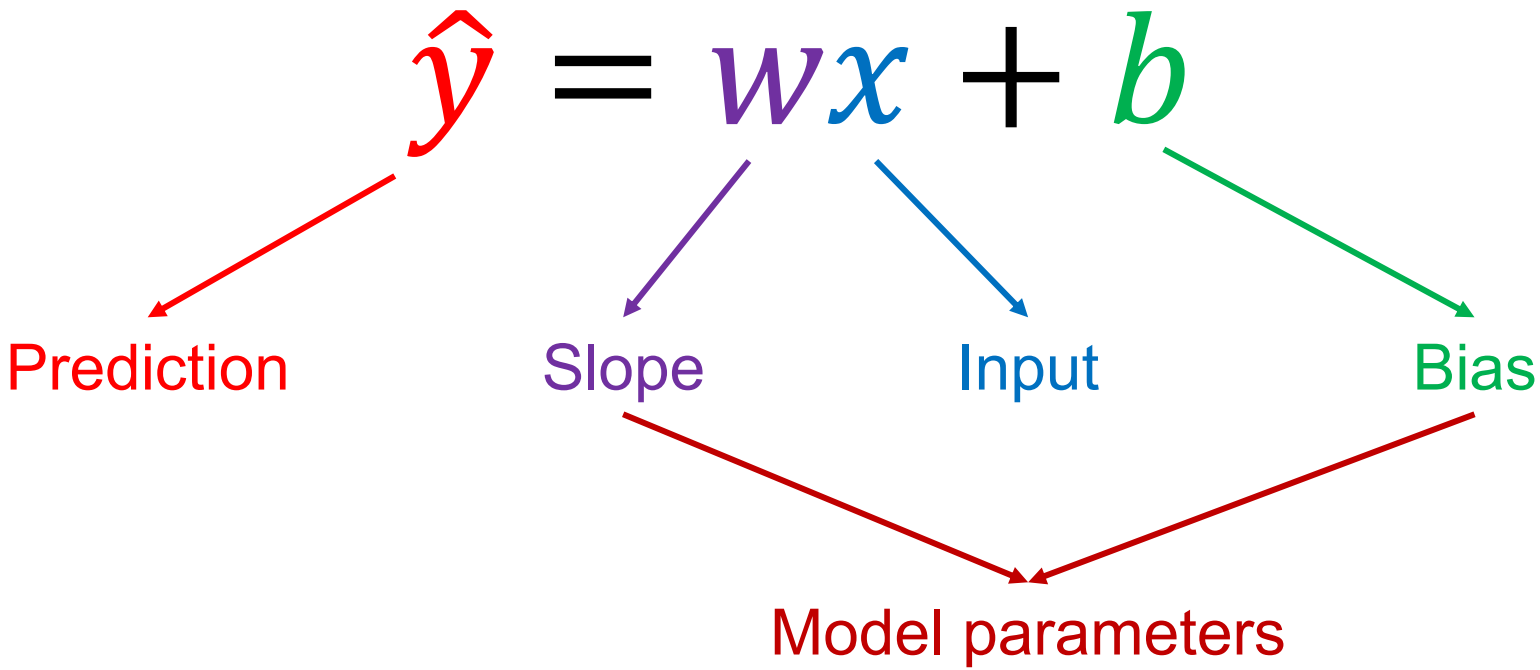
- **Can we describe this relationship with a line?**

- X-axis: house size
- Y-axis: price
- Scatter points

- Goal: find a line that best fits the data

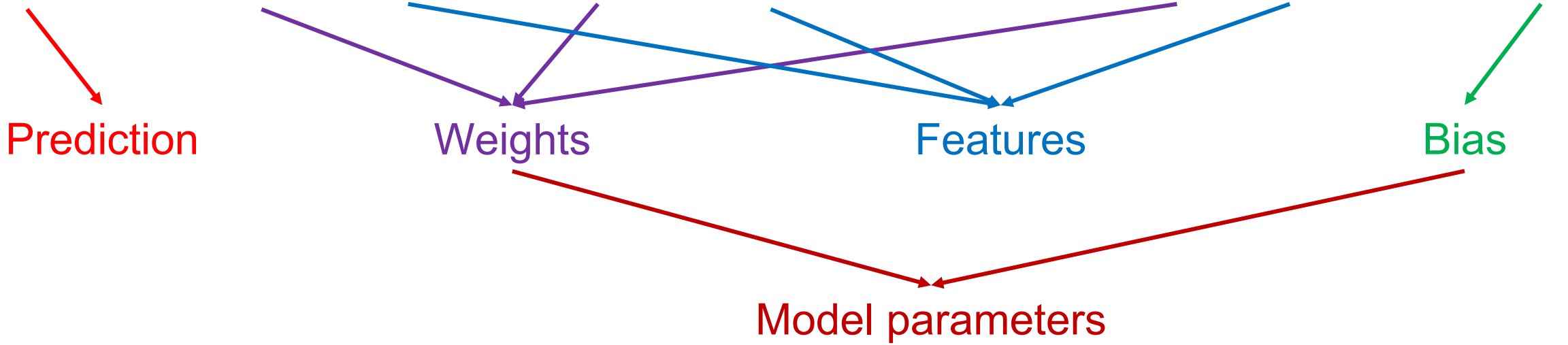


Linear Model



Linear Model (Multiple Features)

$$\hat{y} = w_1 x_1 + w_2 x_2 + \cdots + w_d x_d + b$$

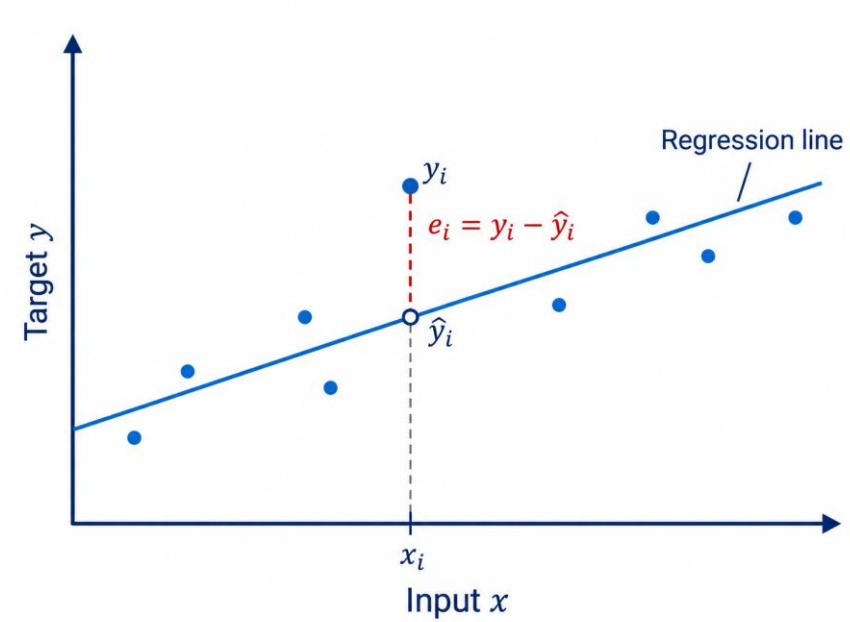


Prediction Error

- Error = difference between target and prediction

- About a data point:

$$\hat{y}_i = wx_i + b, \qquad e_i = y_i - \hat{y}_i$$



Loss Function

- Measures how well the model fits the data
- Converts prediction errors into a single value

- Small loss → better fit
- Large loss → worse fit



- Common examples
  - MSE (regression): squared error
  - MAE (regression): absolute error
  - Cross-Entropy (classification): for class prediction

Loss Function

- Mean Squared Error (MSE)

$$L(w,b) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - (wx_i + b))^2$$

  - Difference → square → average
  - Why square?
    - Prevents positive and negative errors from canceling out
    - Penalizes large errors more

Minimizing the Loss

- Goal: find parameters that minimize the loss

$$w^*, b^* = \arg\min_{w,b} L(w, b)$$

- Learning = finding model parameters that give a small loss

Parameters (w, b)

↓

Prediction $\hat{y}$

↓

Loss $L(w, b)$

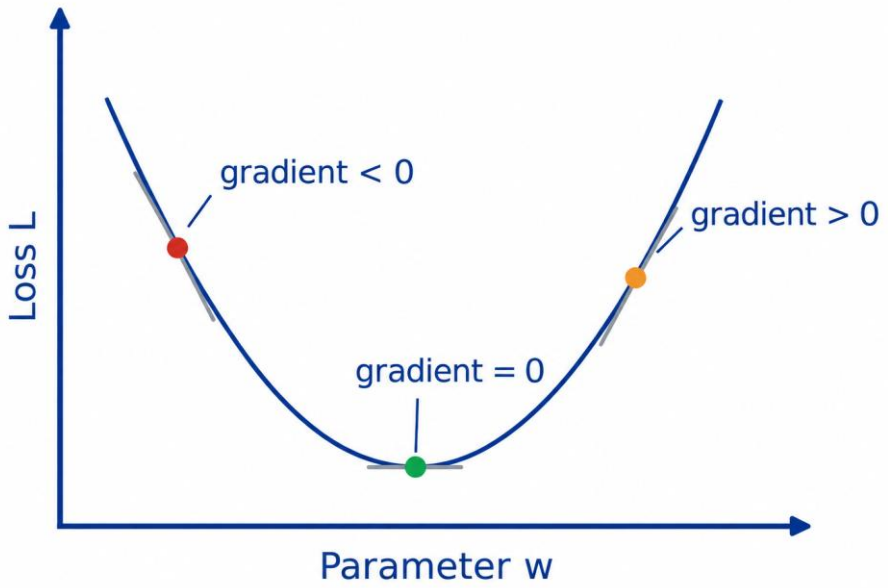
↓

How should we update w, b?

Gradient

- The gradient tells us how the loss changes:

$$\frac{\partial L}{\partial w}$$

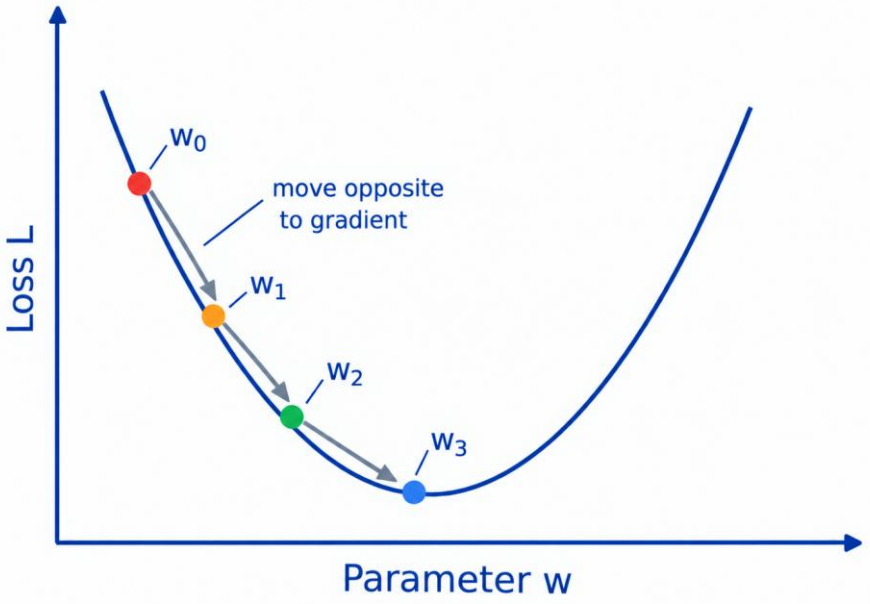


- It points in the direction of increasing loss
- To reduce the loss, move in the opposite direction of the gradient

Gradient Descent

$$w \leftarrow w - \eta \frac{\partial L}{\partial w}$$

- w: model parameter
- $\frac{\partial L}{\partial w}$: gradient
- η: learning rate

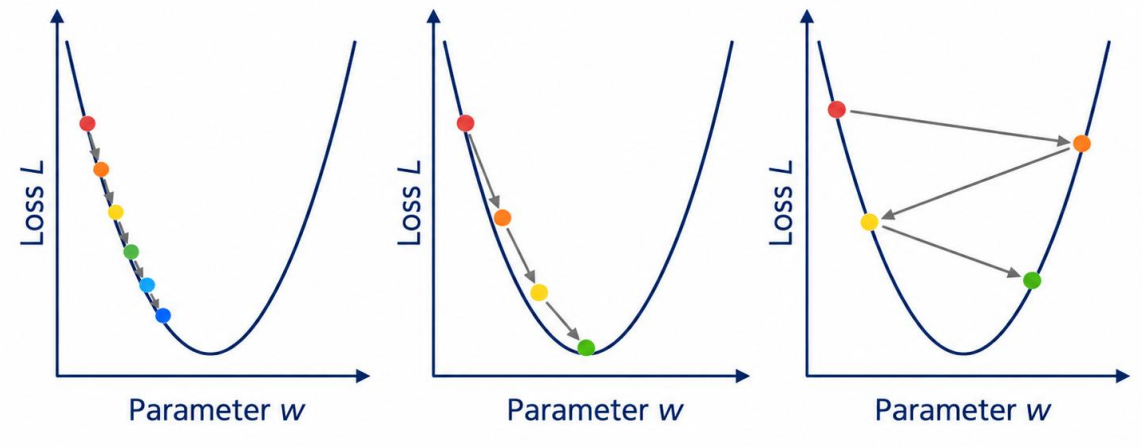


- Move in the opposite direction of the gradient
- Repeat until the loss becomes sufficiently small

Gradient Descent

$$w \leftarrow w - \eta \frac{\partial L}{\partial w}$$

- η: learning rate = step size for each update
  - Too small → slow convergence
  - Appropriate → converges smoothly
  - Too large → overshoots, may diverge



Updating the Linear Model

$$\hat{y} = wx + b$$

- Compute gradients for each parameter
- Update w and b to reduce the loss

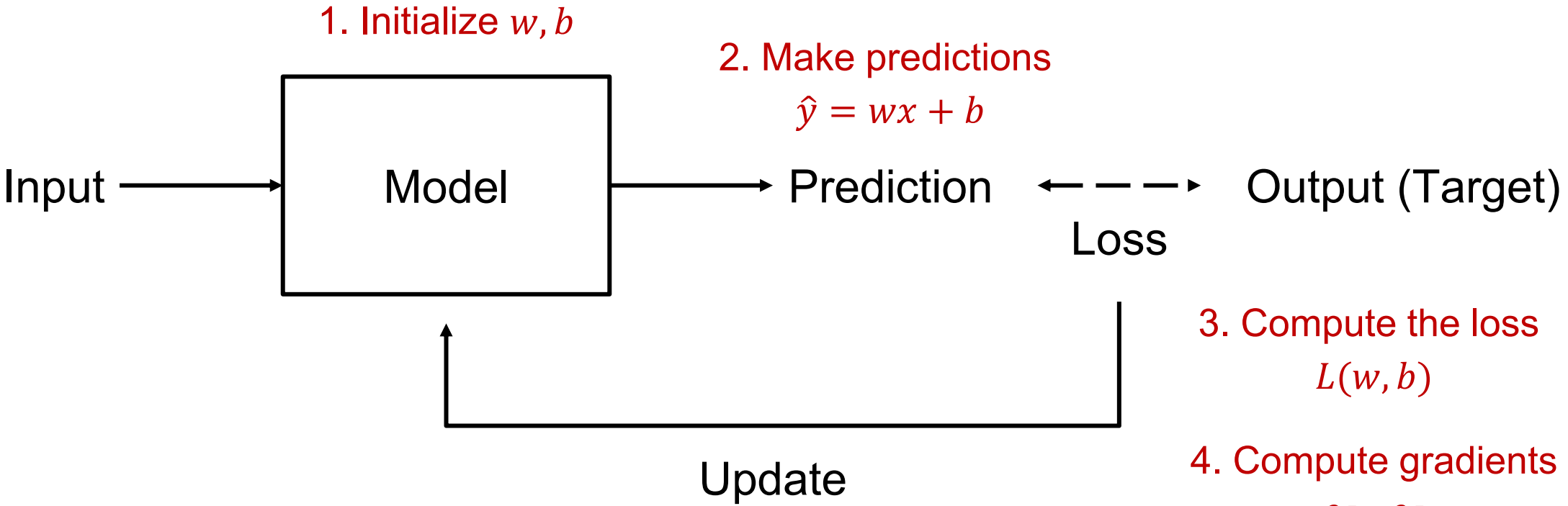
$$w \leftarrow w - \eta \frac{\partial L}{\partial w}$$

$$b \leftarrow b - \eta \frac{\partial L}{\partial b}$$

- For multiple features,

$$w_j \leftarrow w_j - \eta \frac{\partial L}{\partial w_j}, \qquad j = 1, \ldots, d$$

Machine Learning Workflow (revisited)



End-to-End Example

Step 1. Make Predictions

- Given a small dataset:

$$(x_1, y_1) = (1, 3), \qquad (x_2, y_2) = (2, 5), \qquad (x_3, y_3) = (3, 7)$$

- Initial model

$$\hat{y} = wx + b$$
$$w = 1, \qquad b = 1$$

- Predictions

$$\widehat{y_1} = ?, \qquad \widehat{y_2} = ?, \qquad \widehat{y_3} = ?$$

Step 2. Compute the Loss

- Prediction errors

$$e_i = y_i - \hat{y}_i$$

$$e_1 = ?, \qquad e_2 = ?, \qquad e_3 = ?$$

- Mean Squared Error (MSE)

$$L(w, b) = \frac{1}{3}\sum_{i=1}^{3} {e_i}^2 = ?$$

Step 3. Update the Parameters

- Suppose the gradients are:

$$\frac{\partial L}{\partial w} = -9.33, \qquad \frac{\partial L}{\partial b} = -4.00$$

- Learning rate

$$\eta = 0.05$$

- Update w, b

$$w_{new} = w - \eta \frac{\partial L}{\partial w} = ?$$

$$b_{new} = b - \eta \frac{\partial L}{\partial b} = ?$$

- Updated model

$$?$$

Step 4. Check the Improvement

x_i	y_i	Old $\hat{y}_i$	New $\hat{y}_i$
1	3	?	?
2	5	?	?
3	7	?	?

- Before

$$L_{old} = ?$$

- After one update

$$L_{new} = ?$$

- Did the loss decrease?

Next Lecture

- Lecture 3: Classification with Machine Learning
  - Logistic Regression
  - Probabilistic Classification
  - Decision Boundaries
  - Classification Loss Function
  - Binary vs. Multiclass Classification

Korea Aerospace University

Questions?

jiyoonshin@kau.ac.kr