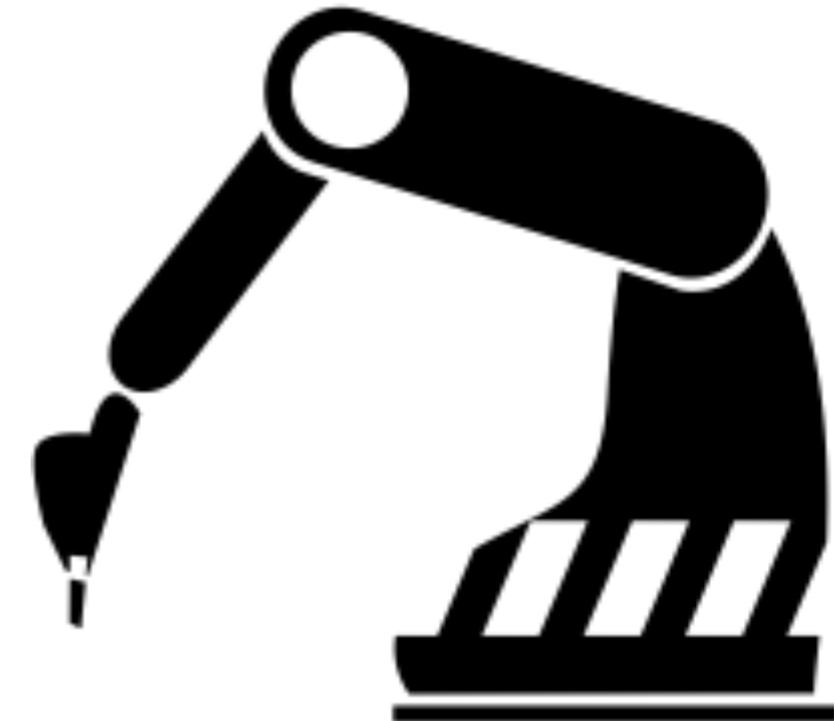
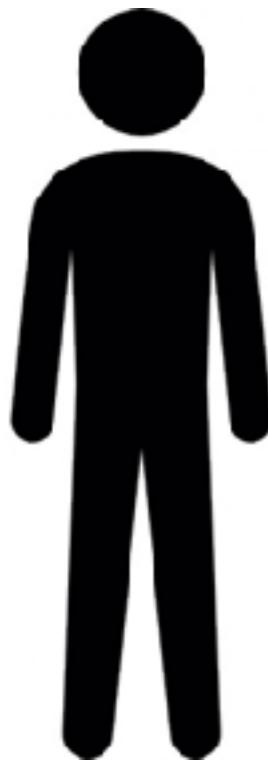


# **Introduction to ML**

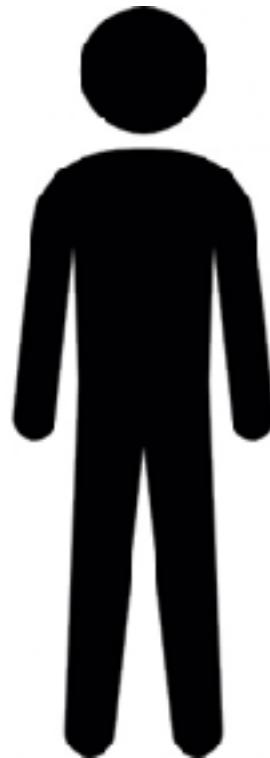
Ermakov Petr

# Machine Learning and Human

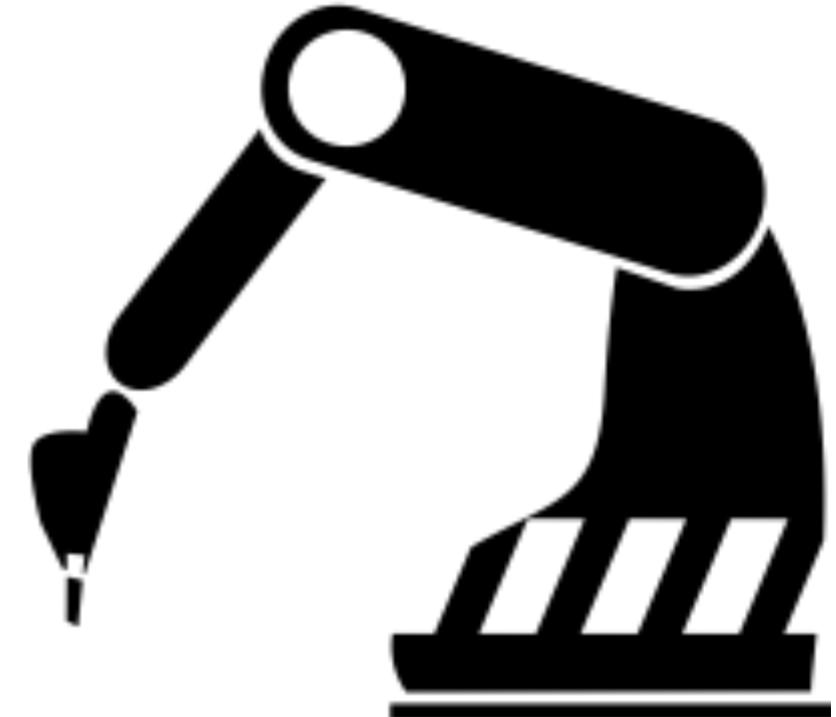


# Machine Learning and Human

**Learn from experience**



**Follow instruction**



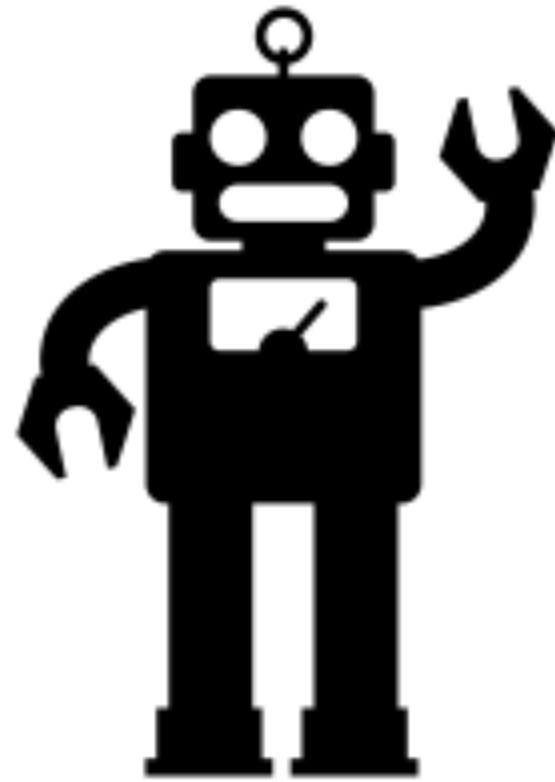
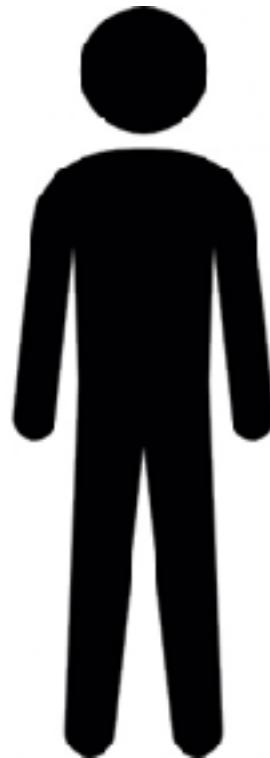




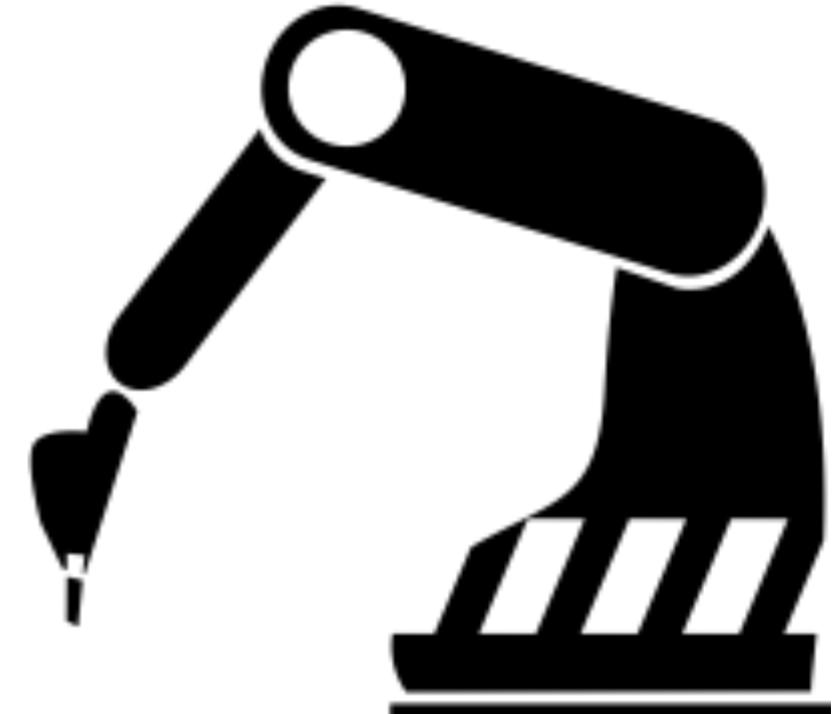


# Machine Learning and Human

**Learn from experience**

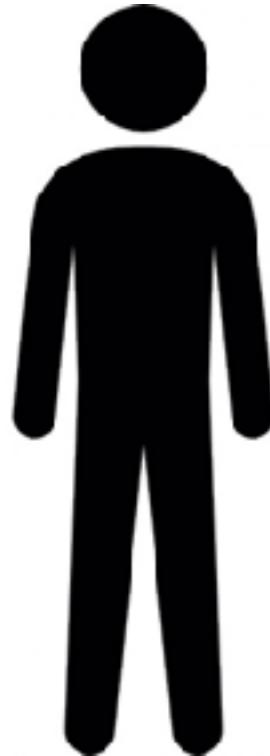


**Follow instruction**

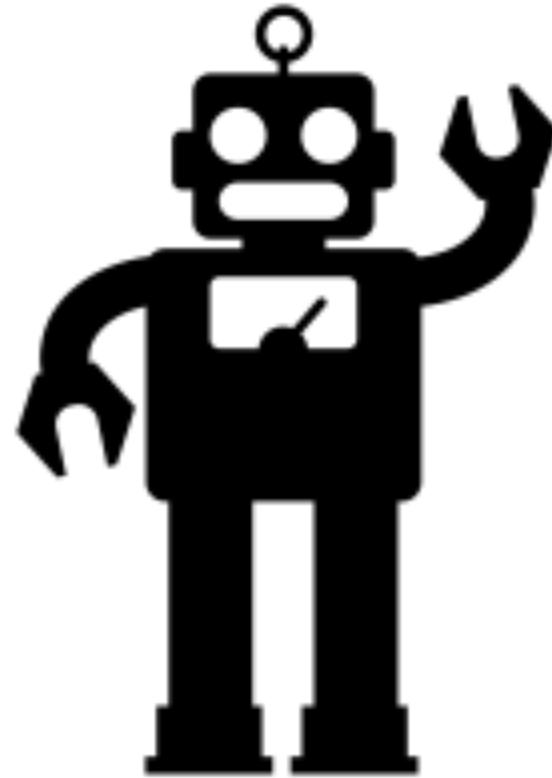


# Machine Learning and Human

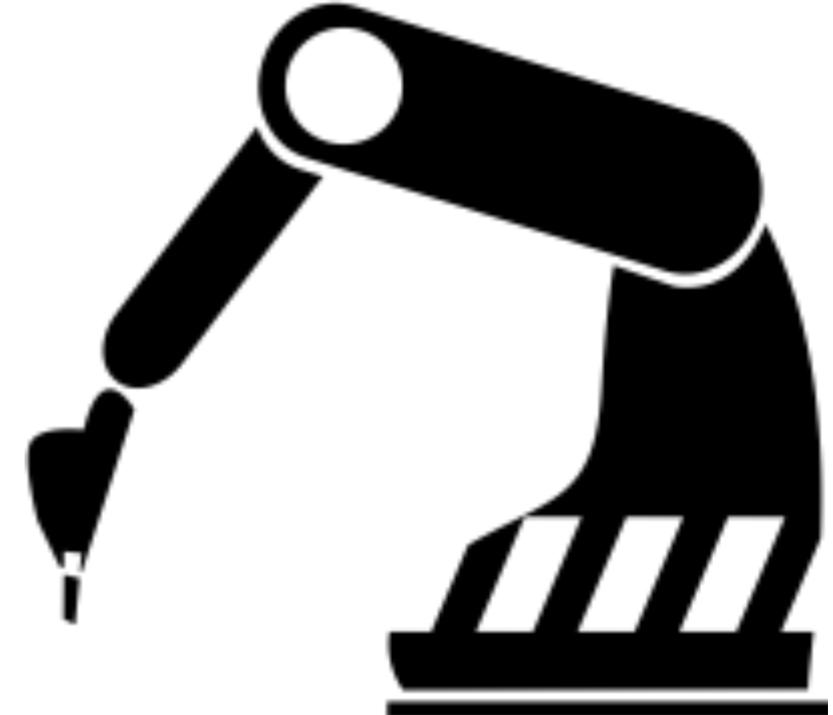
**Learn from experience**



**Learn from experience**



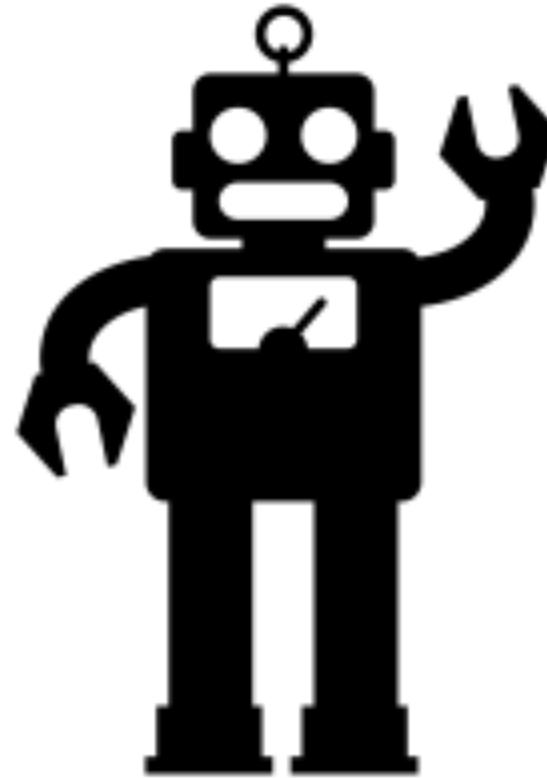
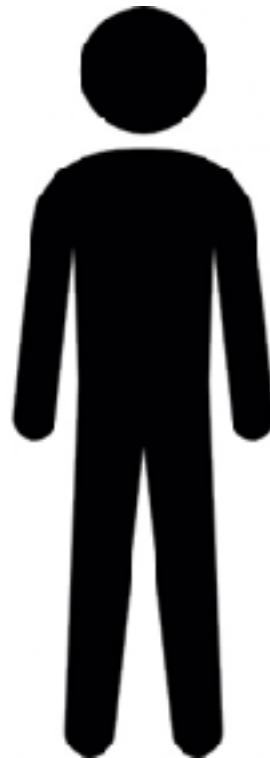
**Follow instruction**



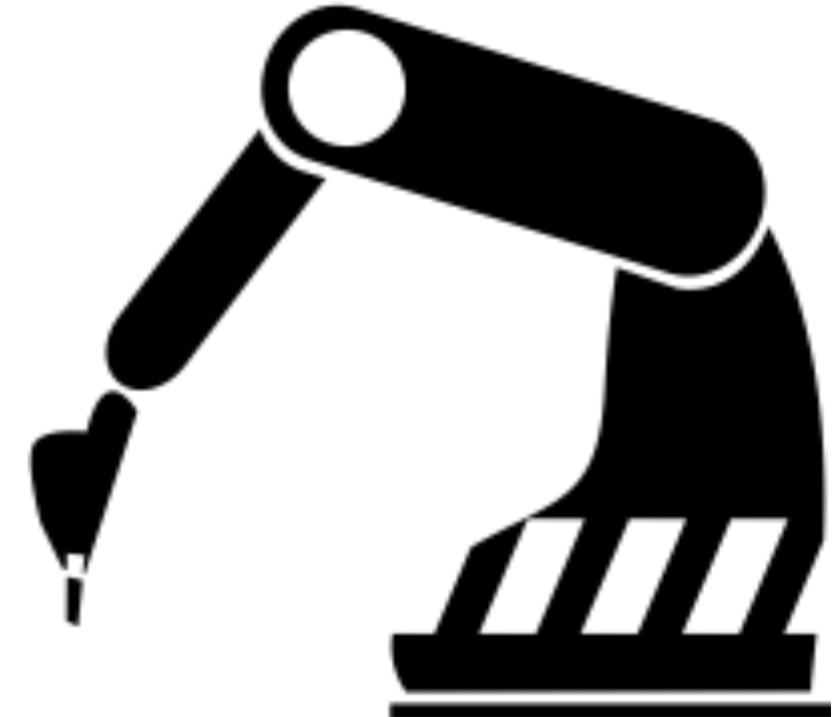
# Machine Learning and Human

data  
~~Learn from experience~~

Learn from experience



Follow instruction



# Machine Learning

## Artificial Intelligence

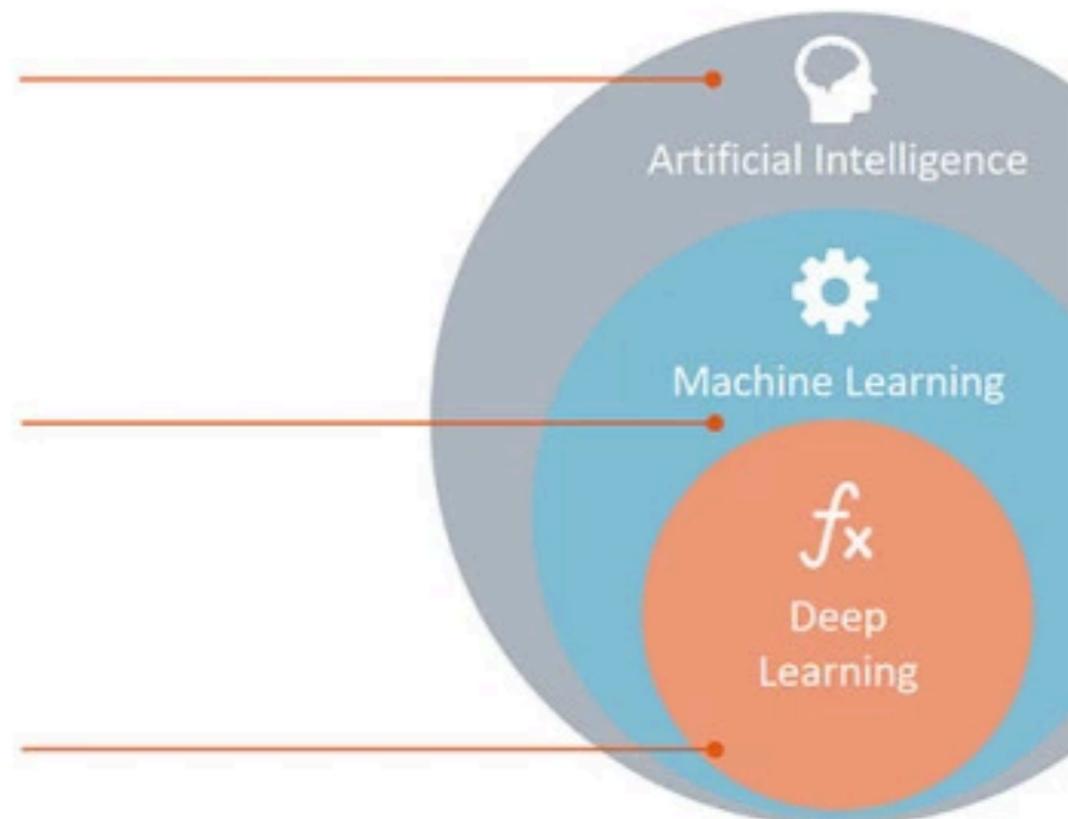
Any technique which enables computers to mimic human behavior.

## Machine Learning

Subset of AI techniques which use statistical methods to enable machines to improve with experiences.

## Deep Learning

Subset of ML which make the computation of multi-layer neural networks feasible.



pic: <https://ingomierswa.com/tag/artificial-intelligence/>



# **Types of Machine Learning**

# Types of Machine Learning

- Supervised learning

# Types of Machine Learning

- Supervised learning
  - Labeled Data

# Types of Machine Learning

- Supervised learning
  - Labeled Data
- Unsupervised learning

# Types of Machine Learning

- Supervised learning
  - Labeled Data
- Unsupervised learning
  - No labels
  - No feedback

# Types of Machine Learning

- Supervised learning
  - Labeled Data
- Unsupervised learning
  - No labels
  - No feedback
- Reinforcement learning

# Types of Machine Learning

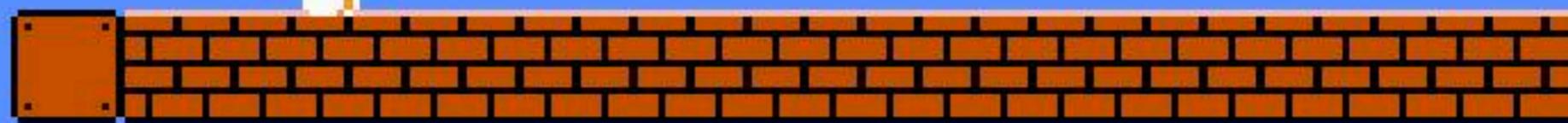
- Supervised learning
  - Labeled Data
- Unsupervised learning
  - No labels
  - No feedback
- Reinforcement learning
  - Learn to react on environment
  - Rewards system

MARIO  
525150

0 x 42

WORLD  
8-2

TIME  
364



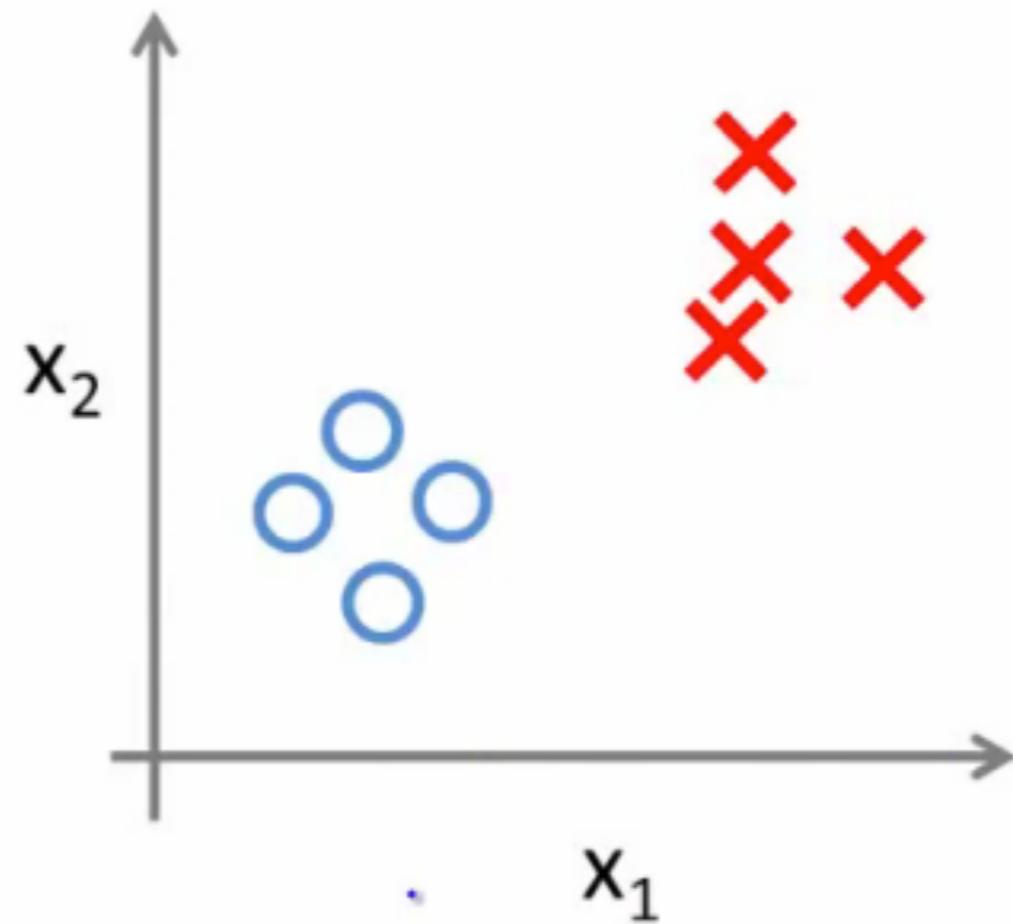
# Supervised learning

- Classification
- Regression
- ... other ...

# Classification

- Binary classification
  - Two classes

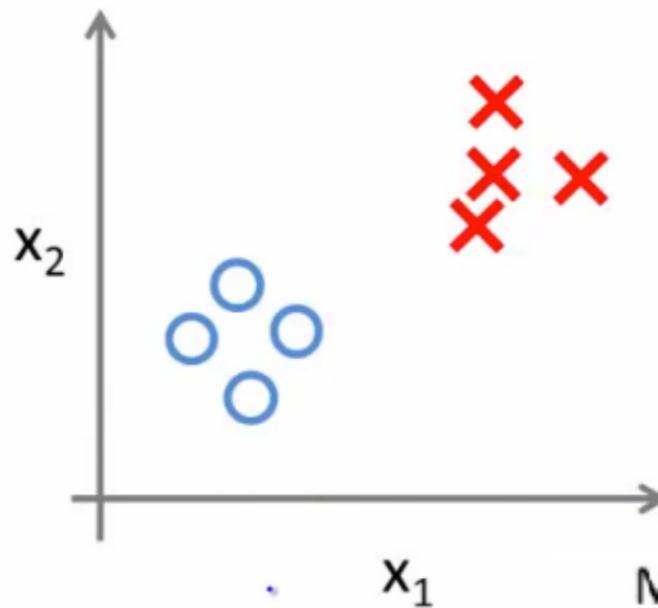
Binary classification:



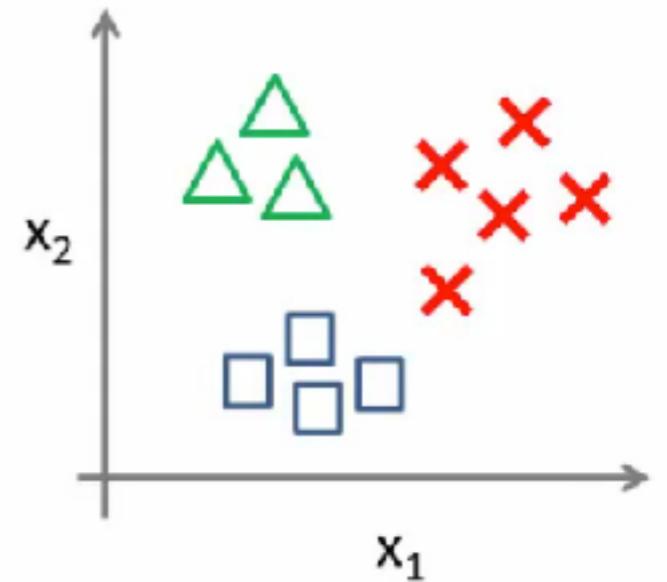
# Classification

- Binary classification
  - Two classes
- Multi-class classification
  - Three or more classes
  - Assigning an object to one of several classes

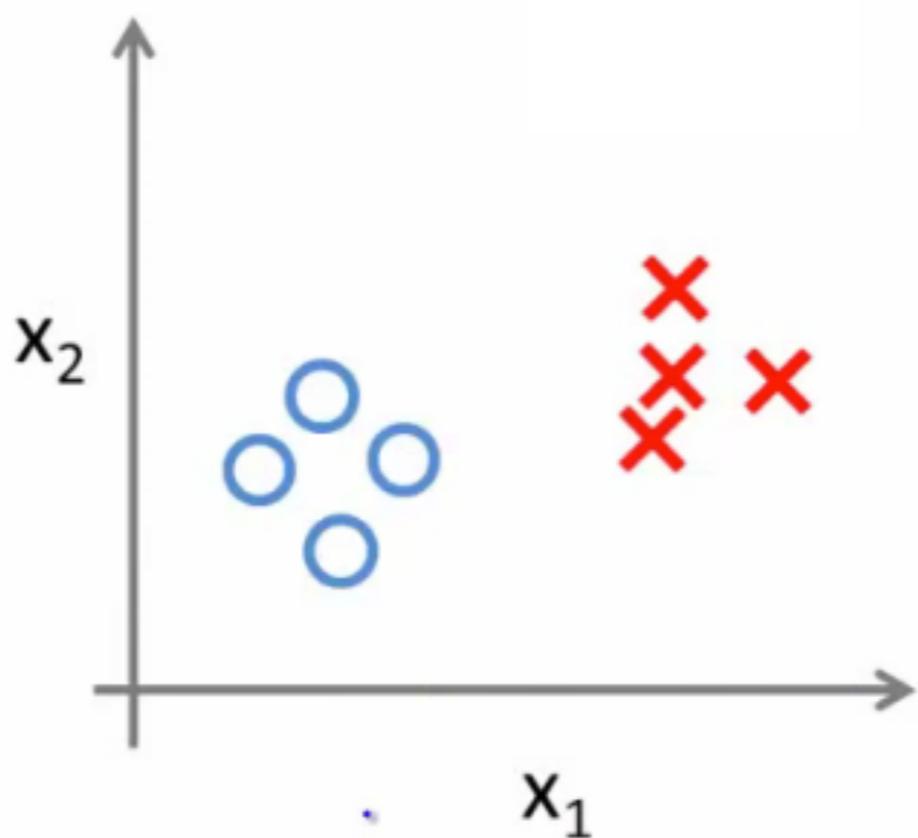
Binary classification:



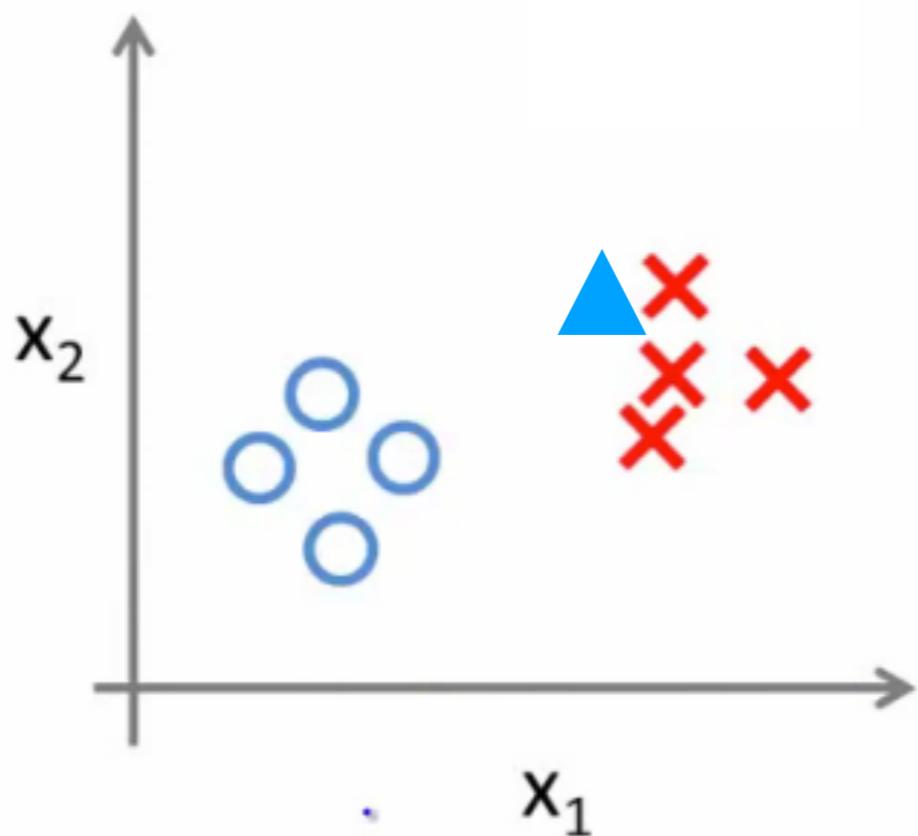
Multi-class classification:



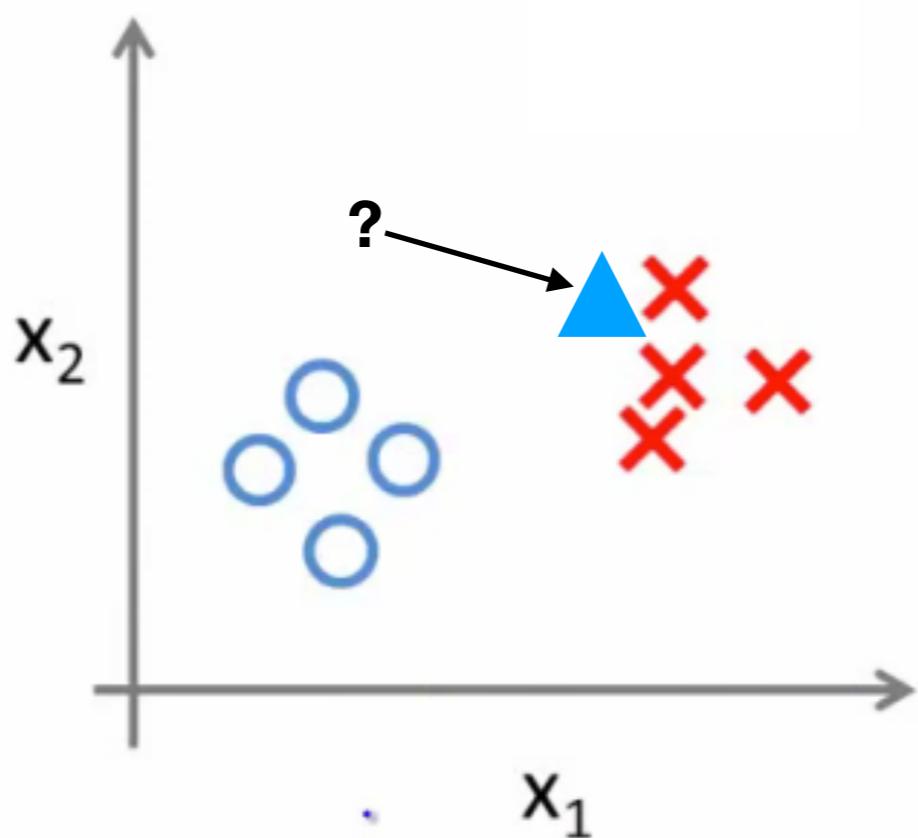
# New element for classification



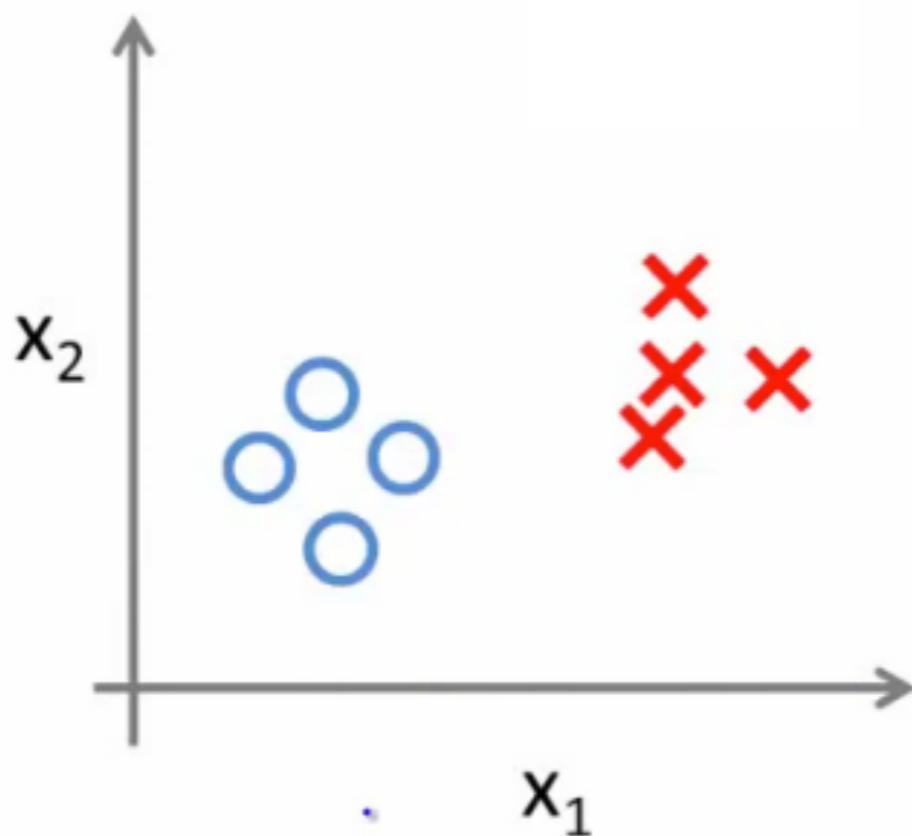
# New element for classification



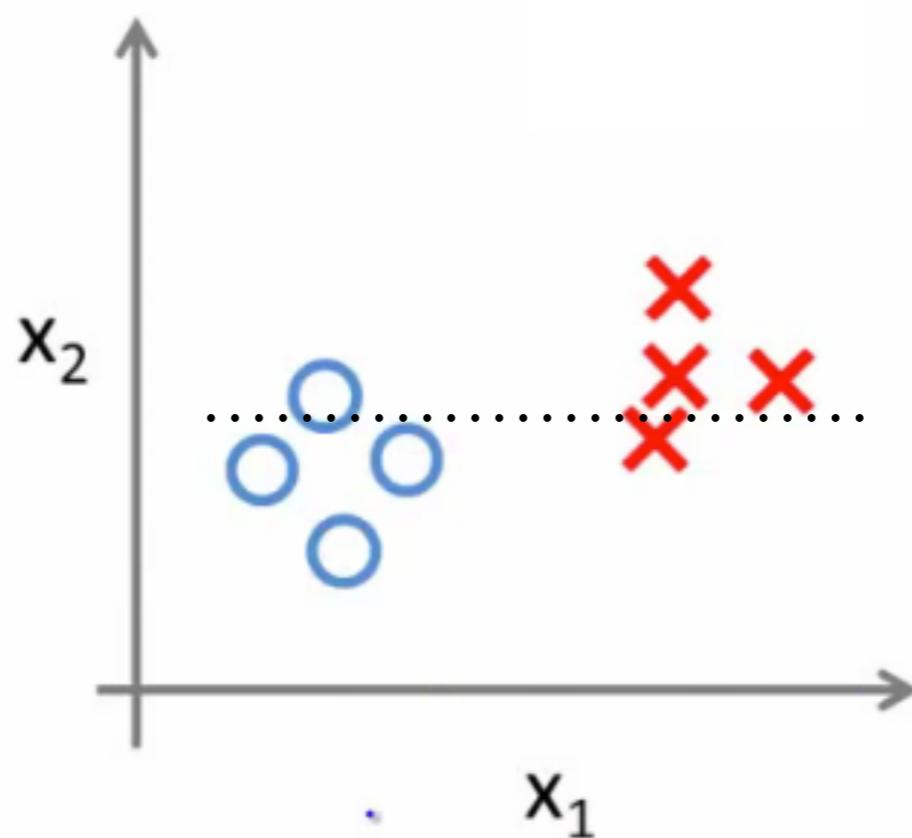
# New element for classification



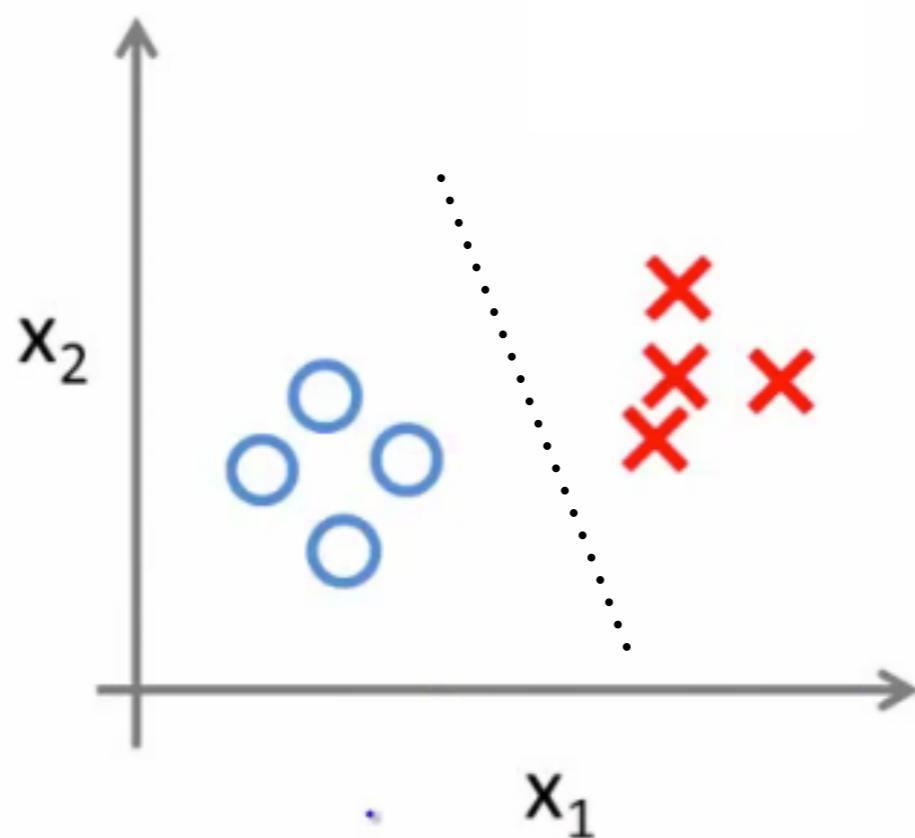
# Build first model



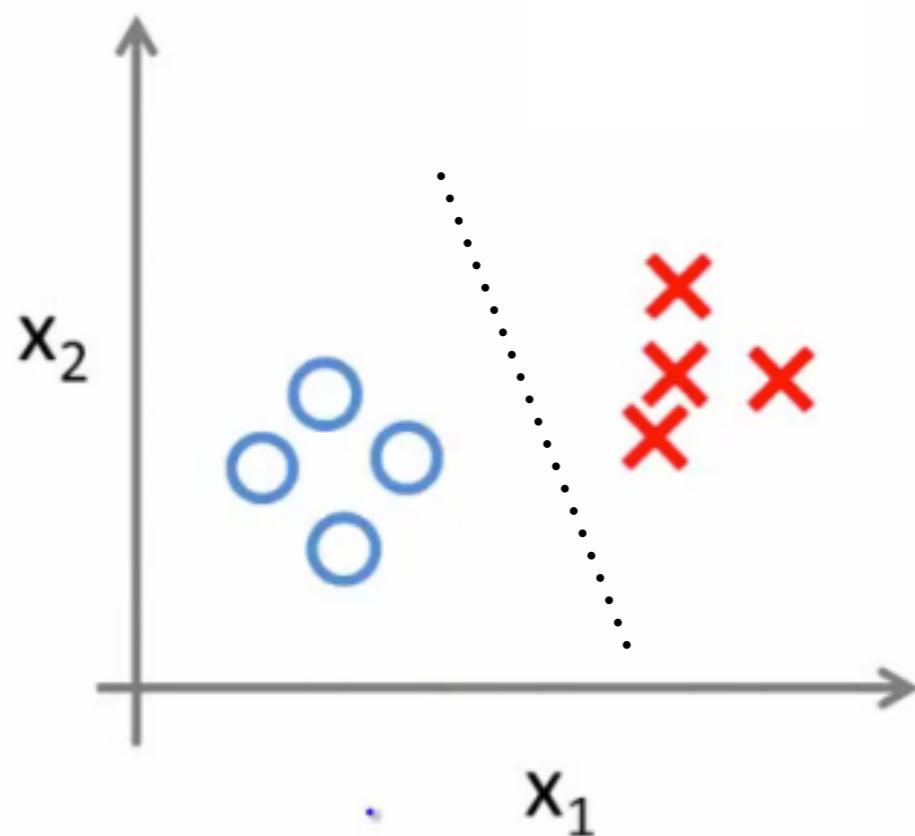
# Build first model



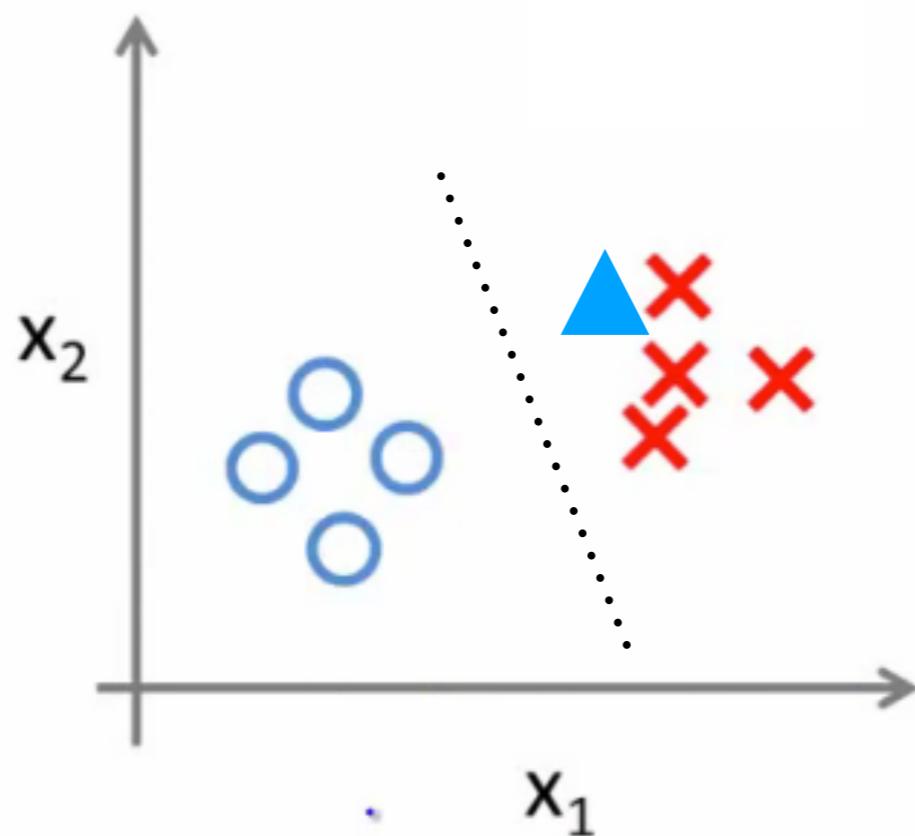
# Build first model



# Build first model



# Build first model



# Classification

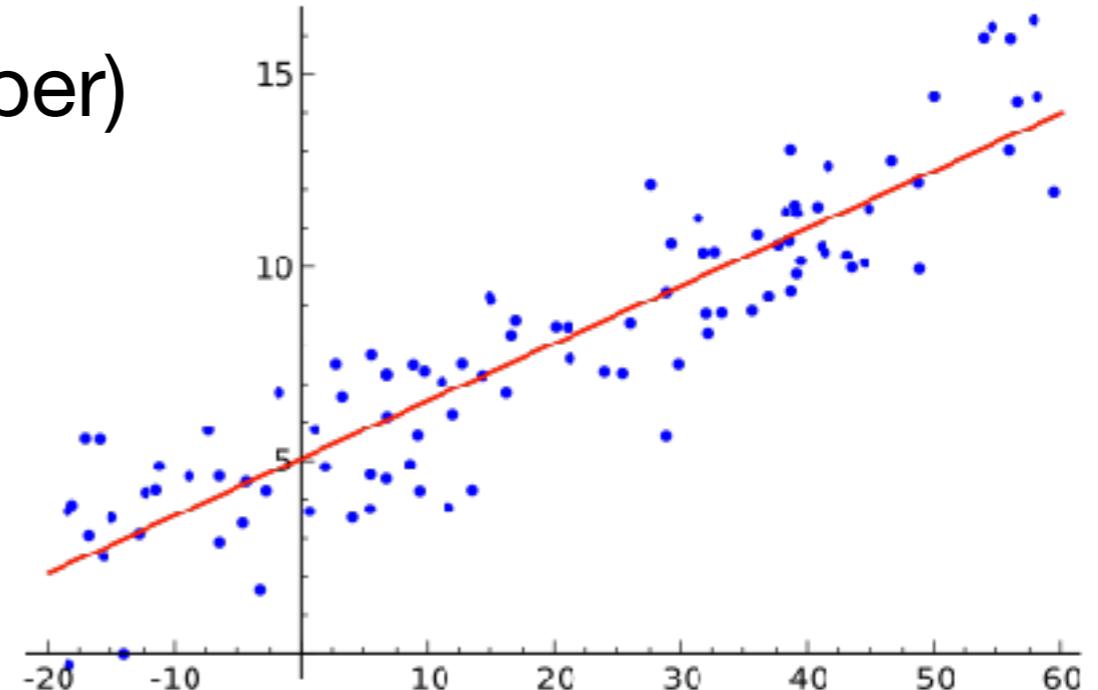
- $X$  - set of object descriptions ( $X_1, X_2$ ) - features
- $Y$  - set of classes (circles and crosses) - target
- $a: X \rightarrow Y$  - ML algorithm

# Titanic

- Which passengers survived the tragedy?
  - X (features):
    - Ticket class
    - Sex
    - Port of Embarkation
    - Ticket number
    - Age in years, ...
  - Y (target):
    - Survival
- 0 = No, 1 = Yes

# Regression

- If for **classification** we build a:  $X \rightarrow Y$   
when  $Y = \{0, 1\}$  or  $Y = \{\text{class1}, \text{class2}, \text{class3}, \dots\}$
- **Regression:**  $Y \in \mathbb{R}$  (real number)
  - Price
  - Age
  - Temperatura



# Regression



# Regression

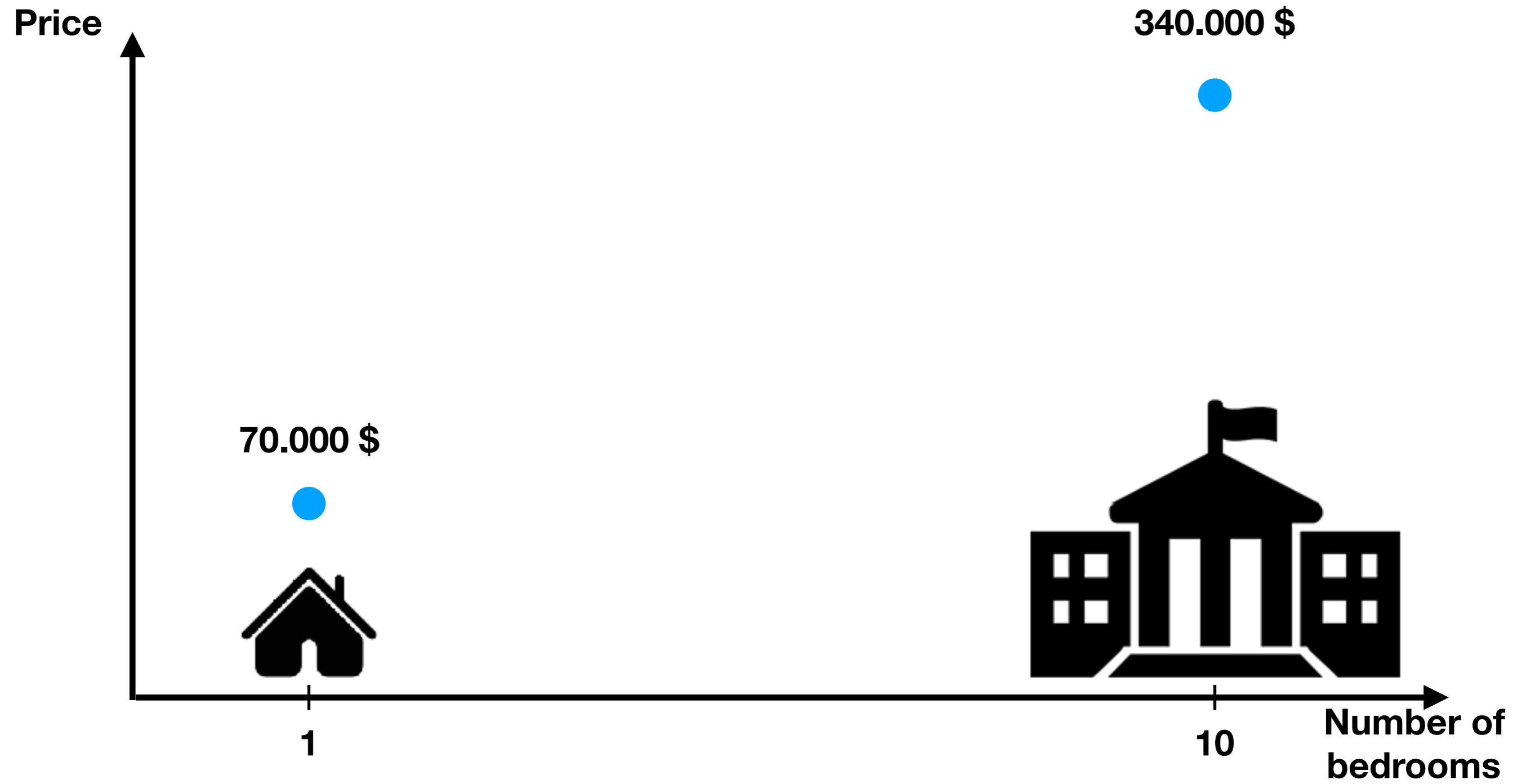
Price



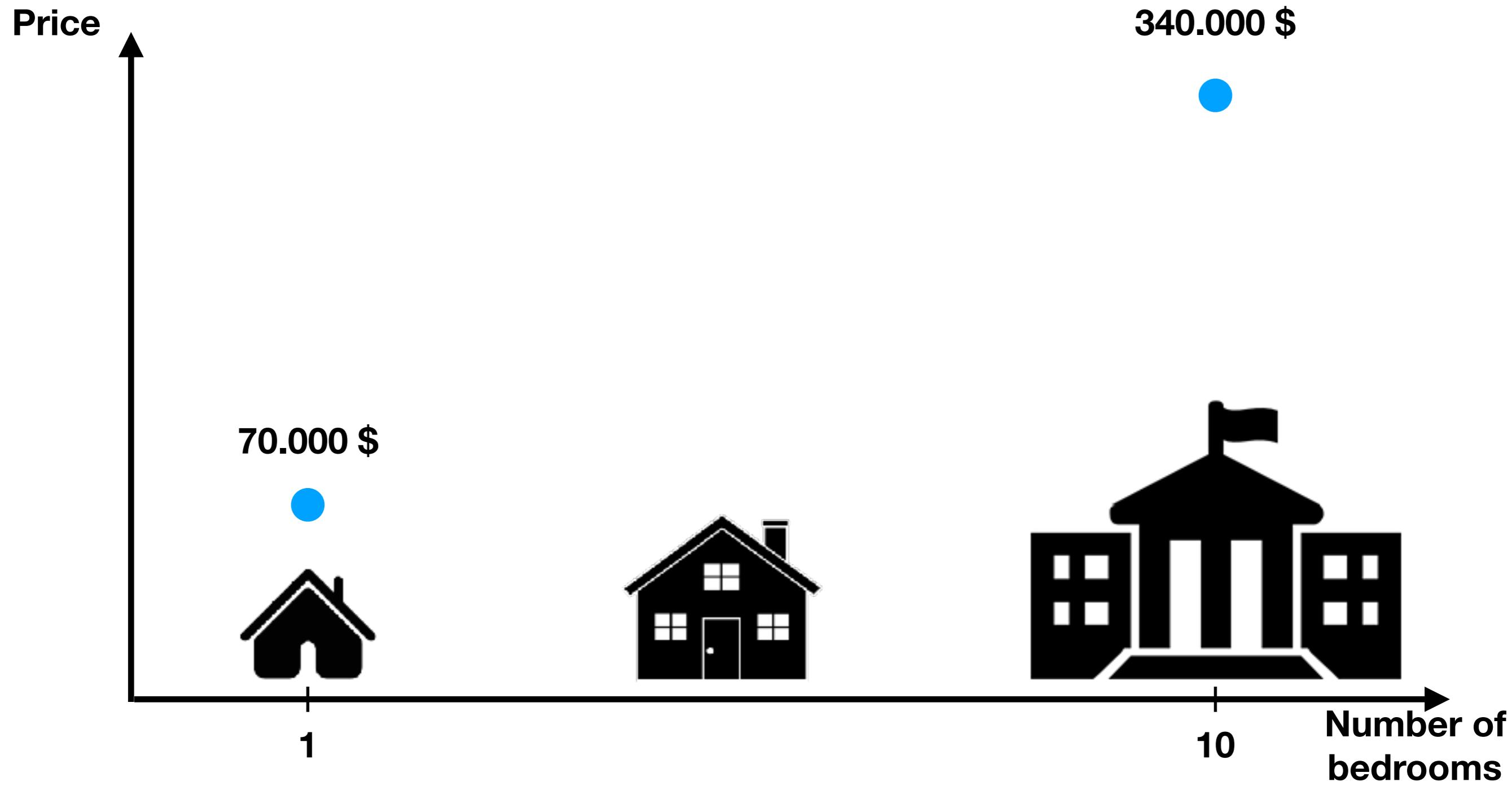
# Regression



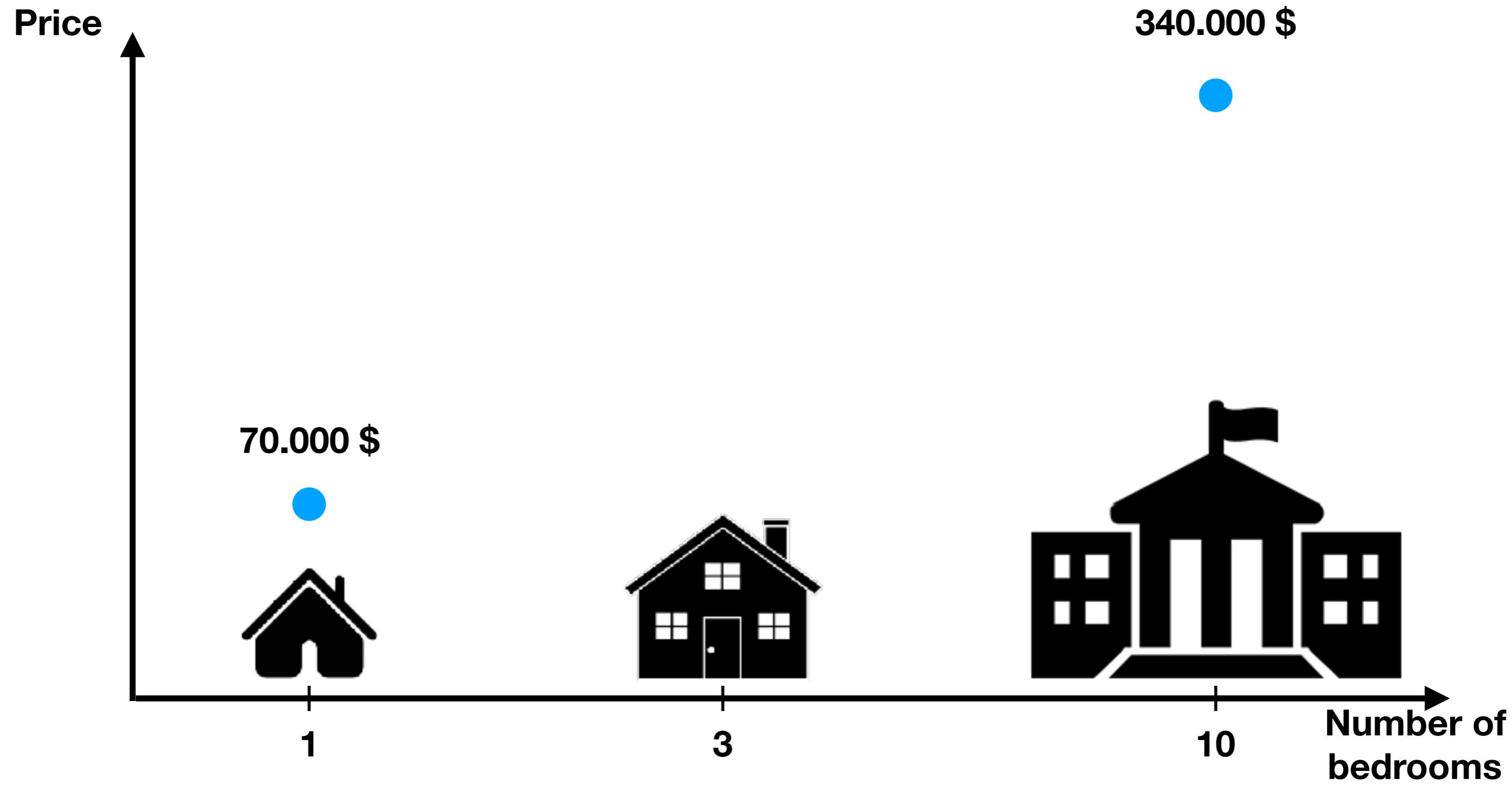
# Regression



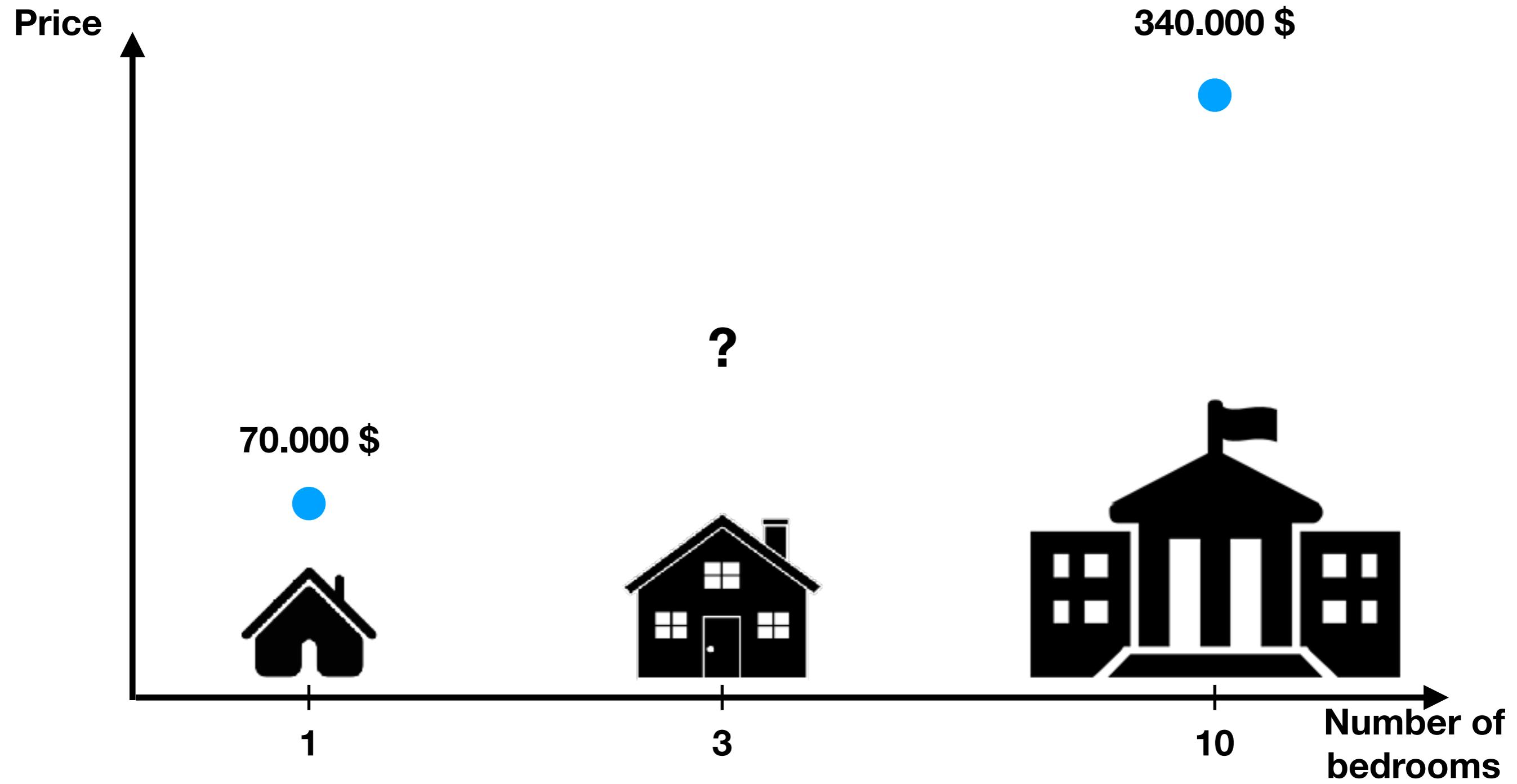
# Regression



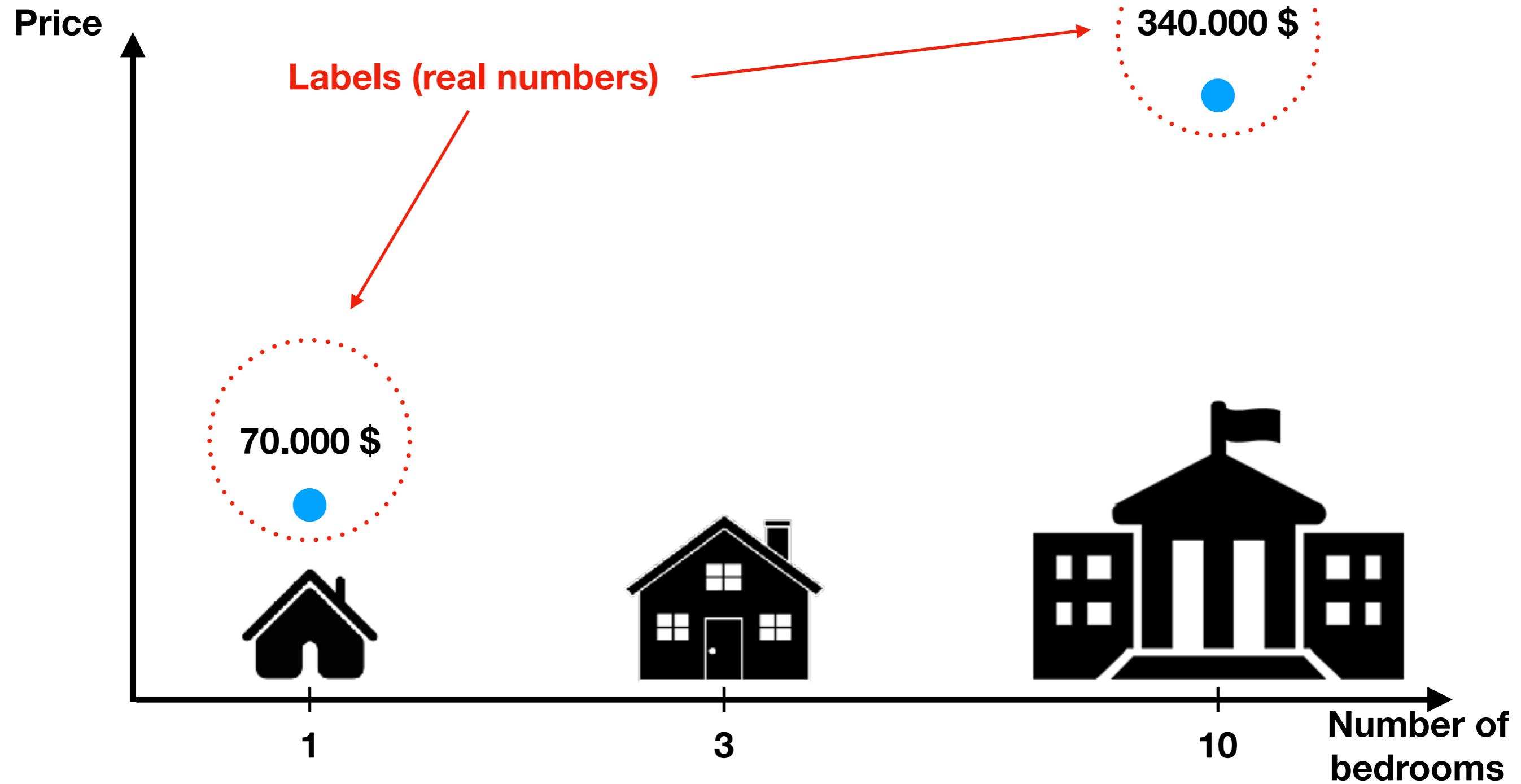
# Regression



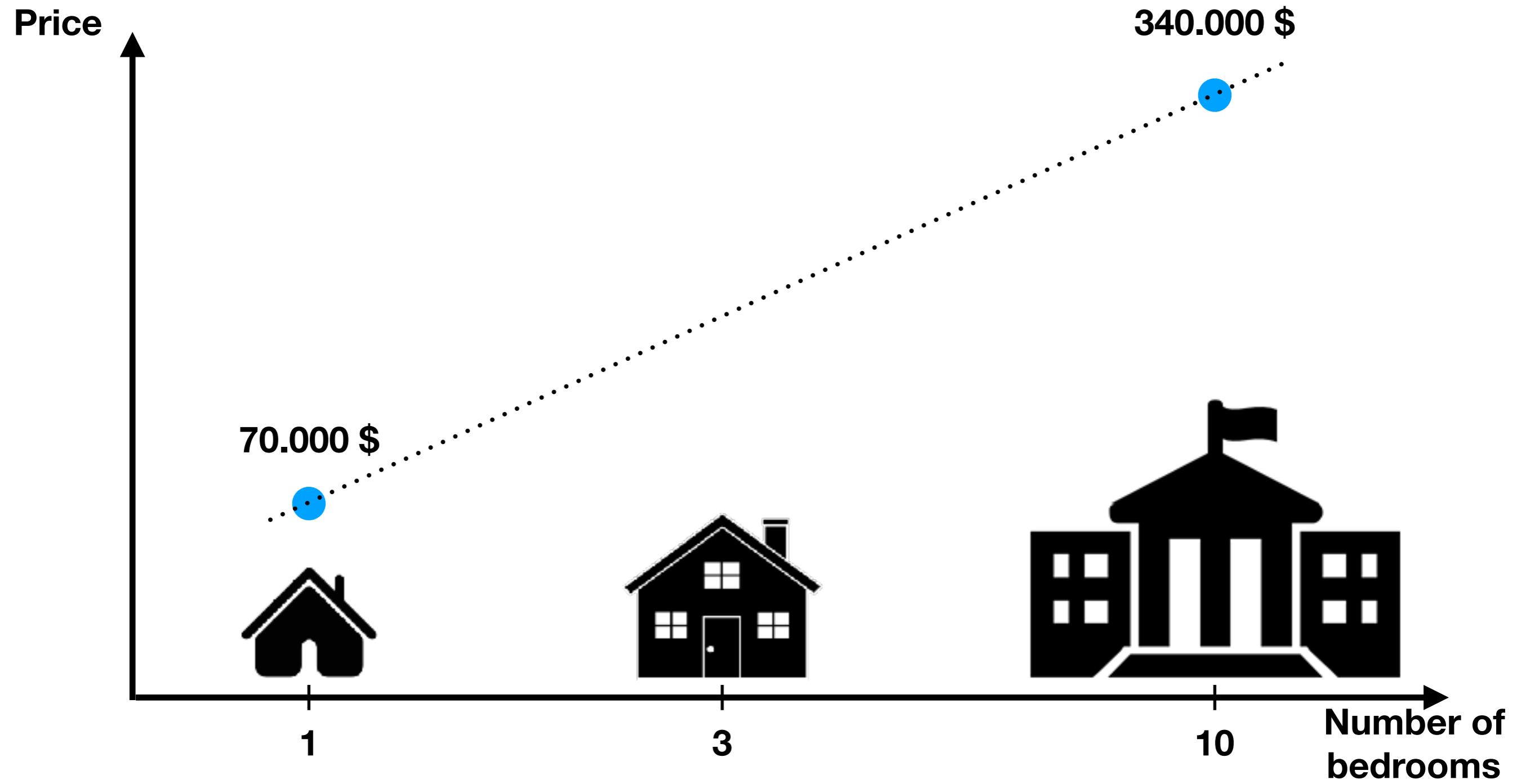
# Regression



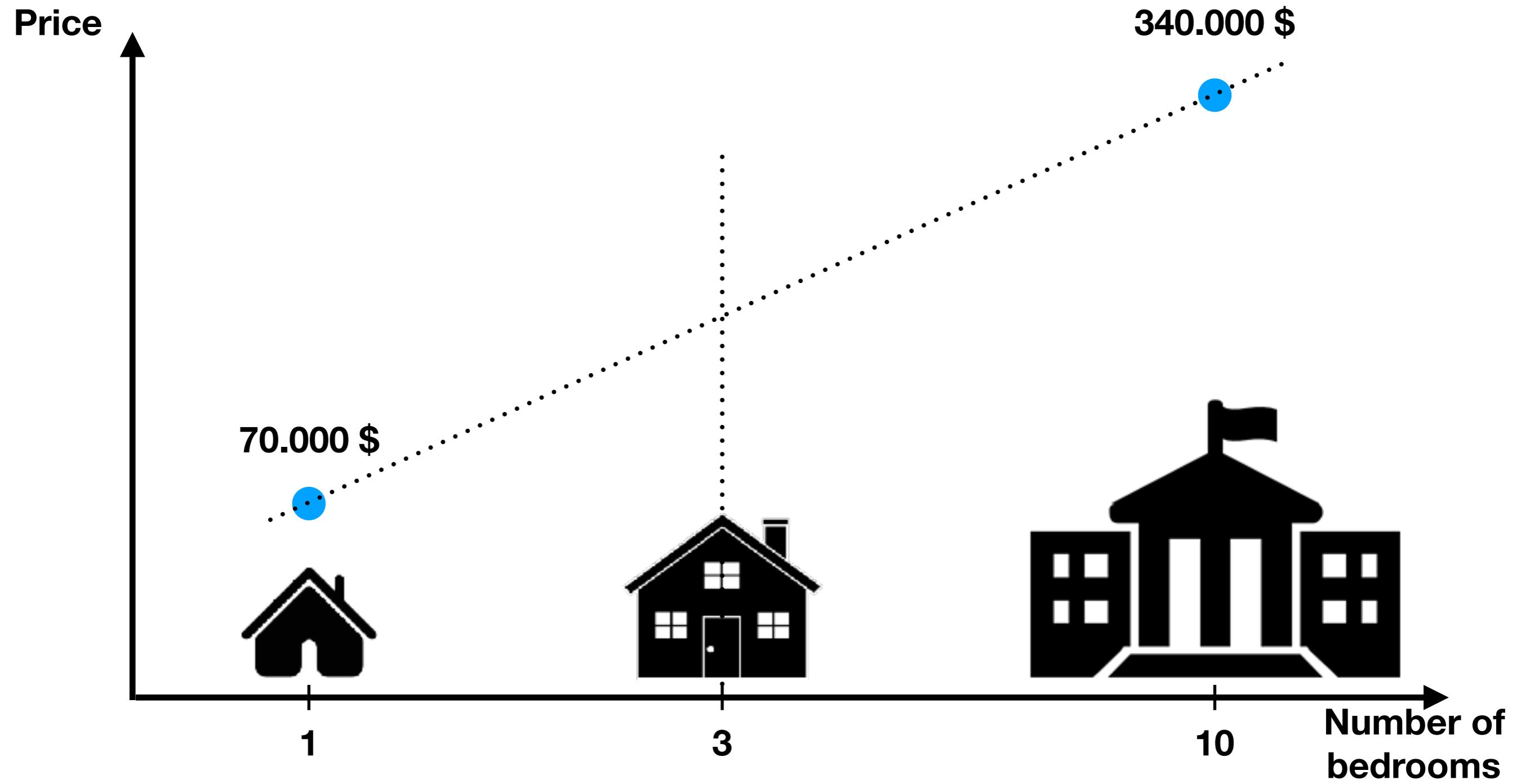
# Regression



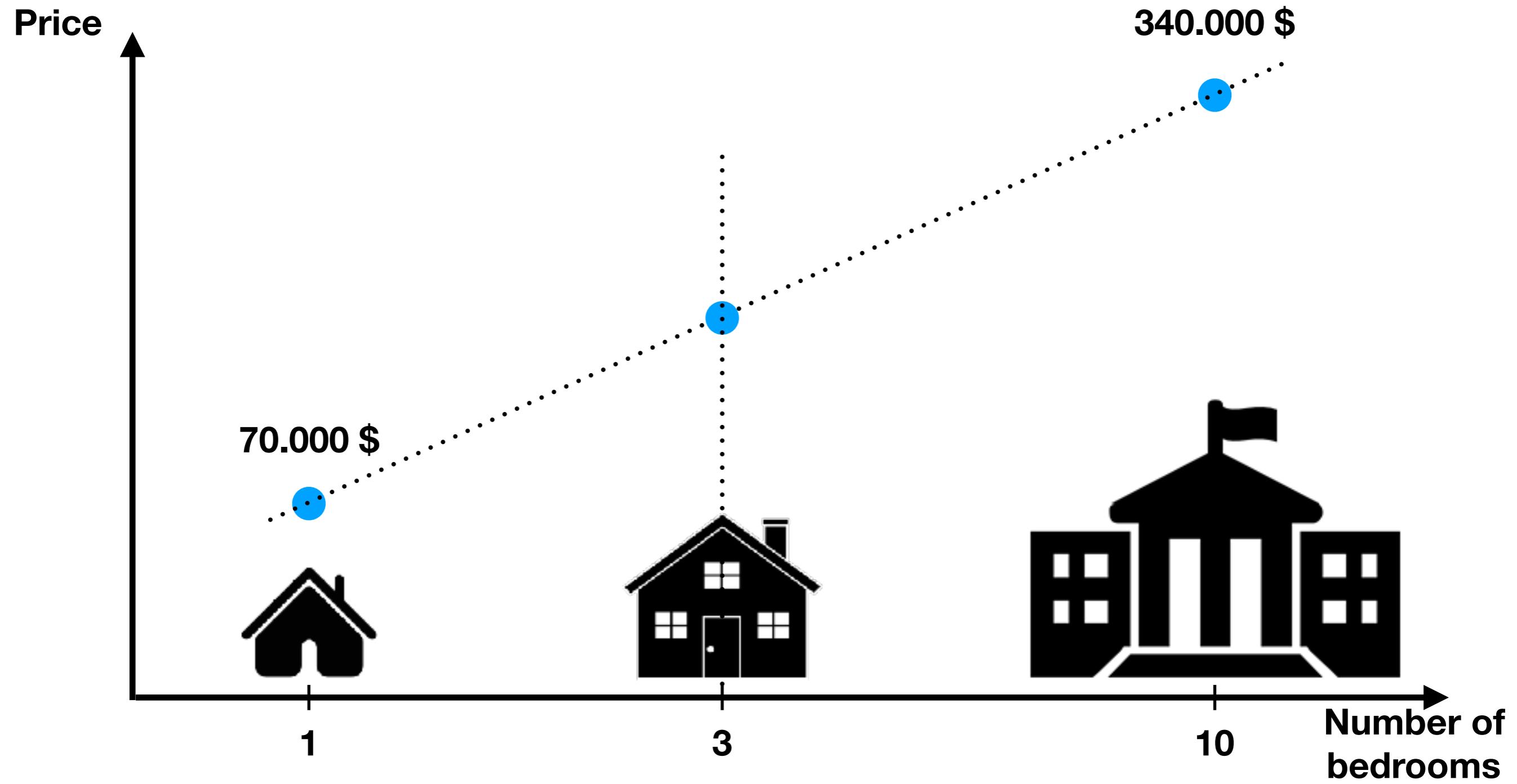
# Regression



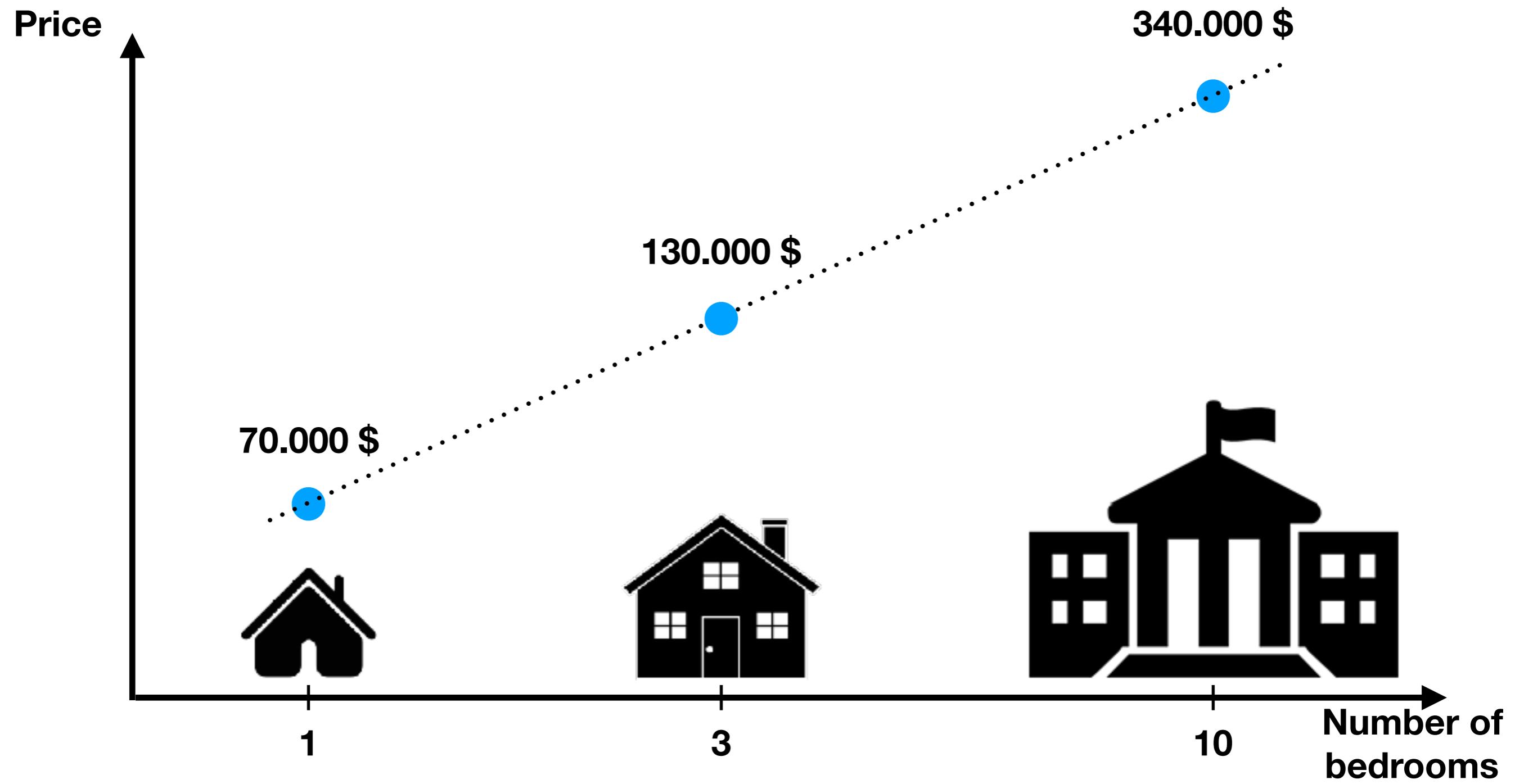
# Regression



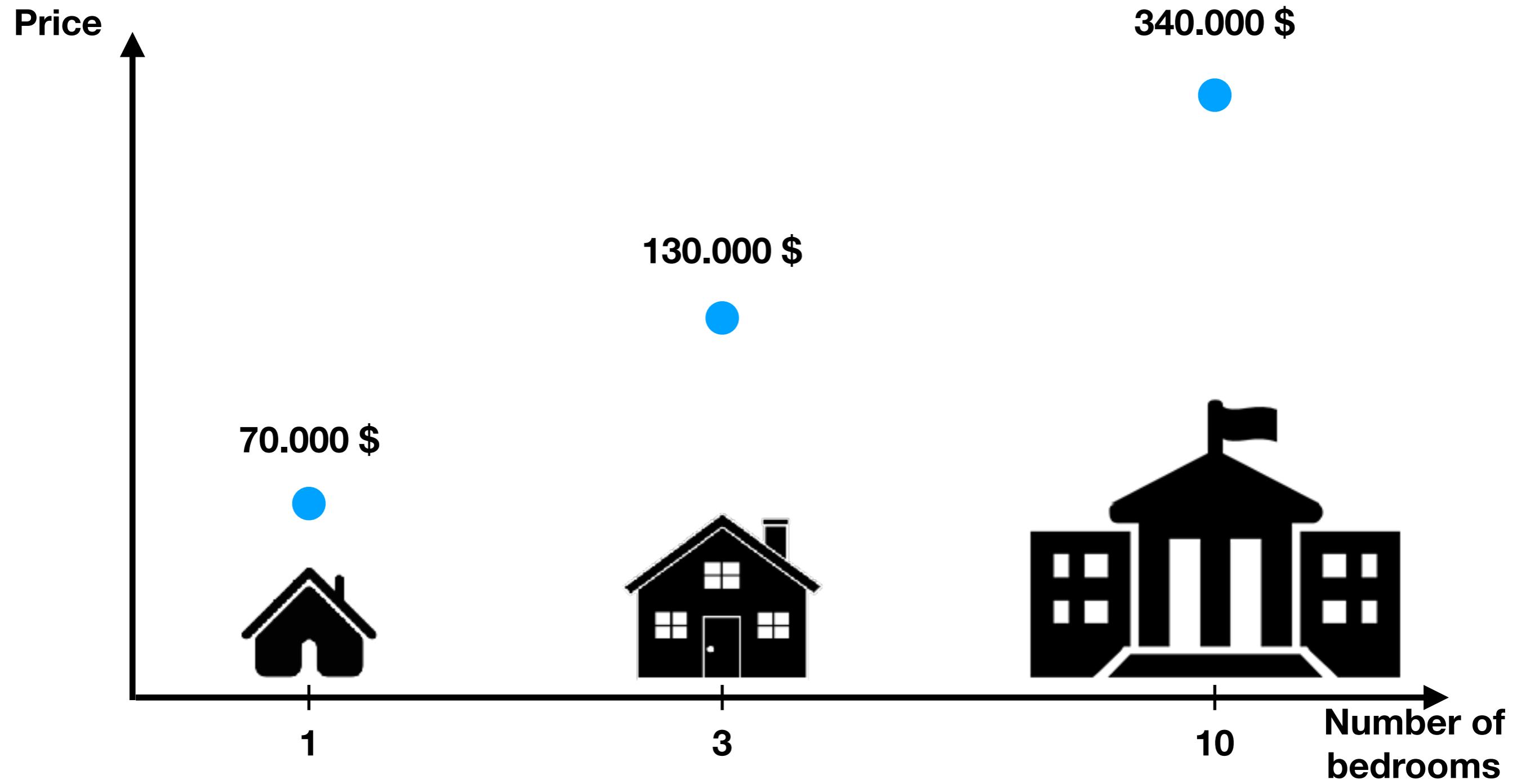
# Regression



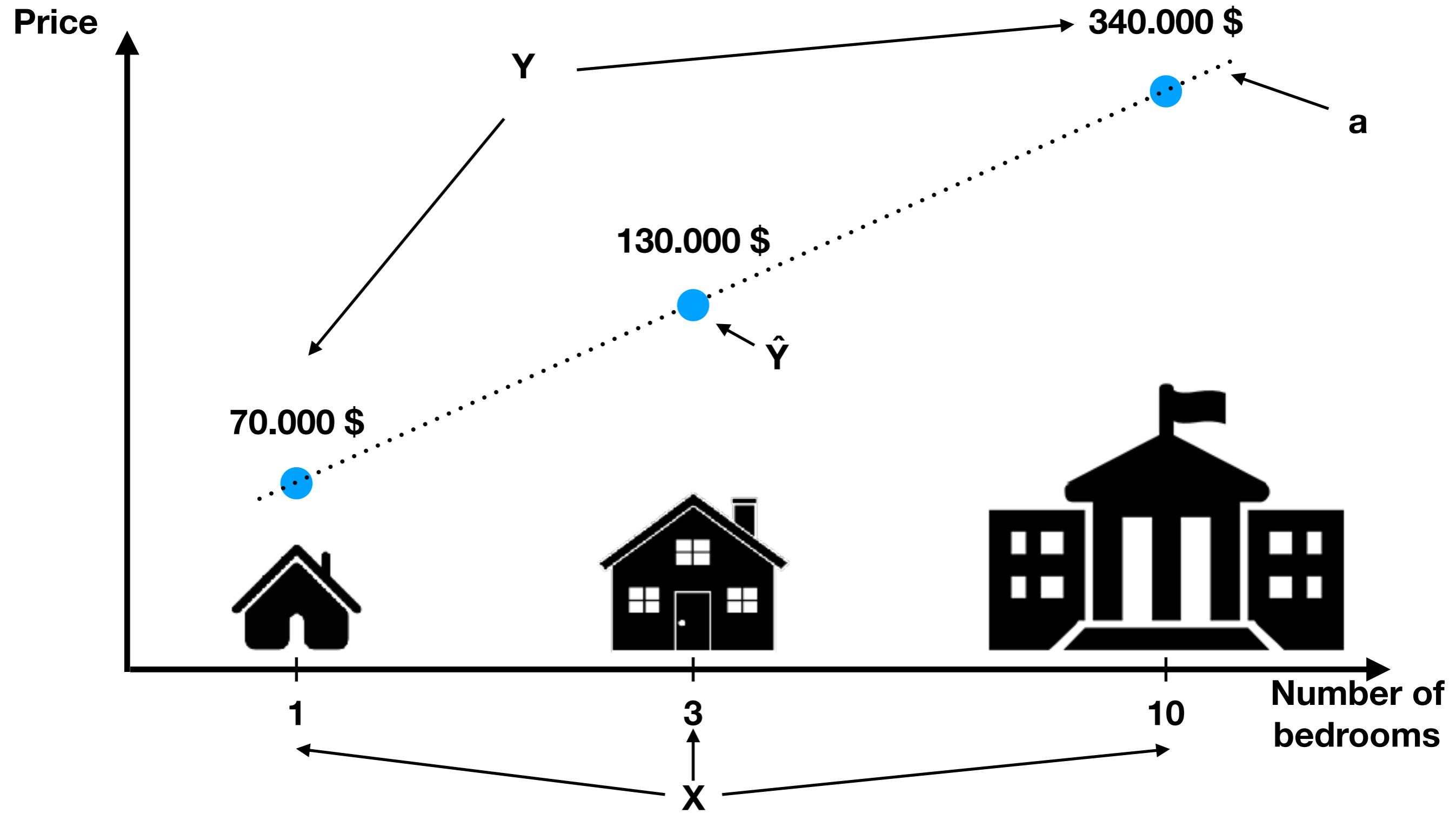
# Regression



# Regression

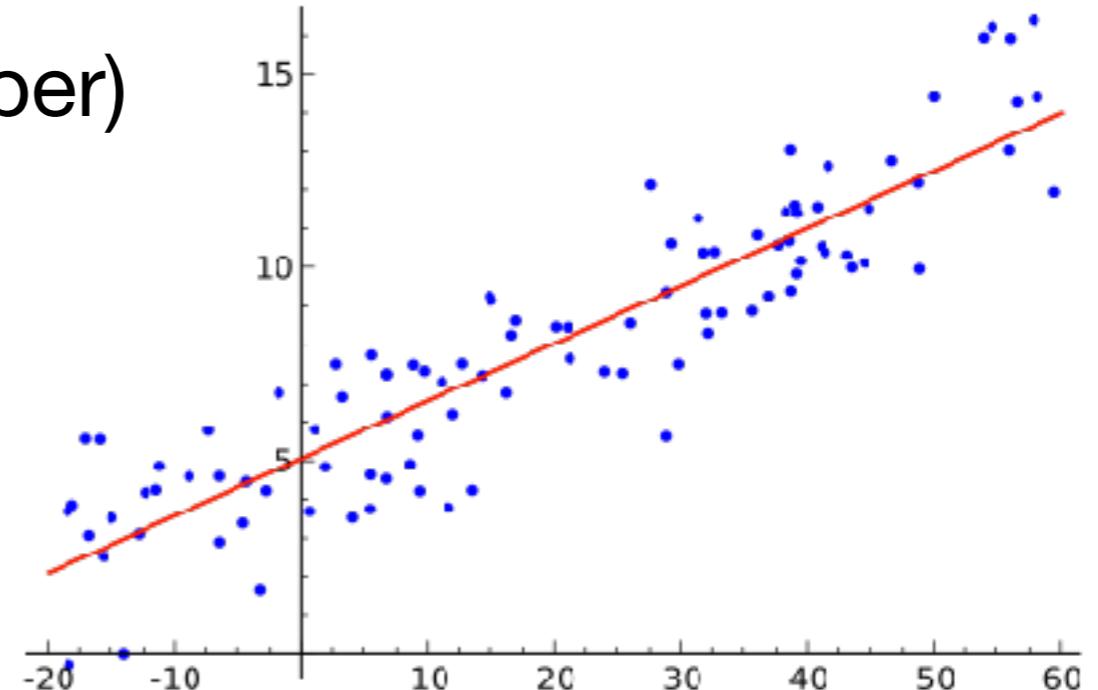


# Regression



# Regression

- If for **classification** we build a:  $X \rightarrow Y$   
when  $Y = \{0, 1\}$  or  $Y = \{\text{class1}, \text{class2}, \text{class3}, \dots\}$
- **Regression:**  $Y \in \mathbb{R}$  (real number)
  - Price
  - Age
  - Temperatura



# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	???

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	???

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

$\hat{Y}$  = *f(total area, number of bedrooms, age of the house)*

$$\hat{Y} = f(X_1, X_2, X_3)$$

$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	???

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

$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$

$$\hat{Y} = f(X_1, X_2, X_3)$$

$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	???

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

$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$

$$\hat{Y} = f(X_1, X_2, X_3)$$

$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

Need find:  $a, b, c, d$

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	???

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

$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$

$$\hat{Y} = f(X_1, X_2, X_3)$$

$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

Need find:  $a, b, c, d$

If we find:  $Y1 (a1, b1, c1, d1)$  and  $Y2 (a2, b2, c2, d2)$

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	???

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

$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$

$$\hat{Y} = f(X_1, X_2, X_3)$$

$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

Need find:  $a, b, c, d$

If we find:  $\mathbf{Y1}$  ( $a_1, b_1, c_1, d_1$ ) and  $\mathbf{Y2}$  ( $a_2, b_2, c_2, d_2$ )

But what's better?  $\mathbf{Y1}$  or  $\mathbf{Y2}$

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	???

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

$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$

$$\hat{Y} = f(X_1, X_2, X_3)$$

$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

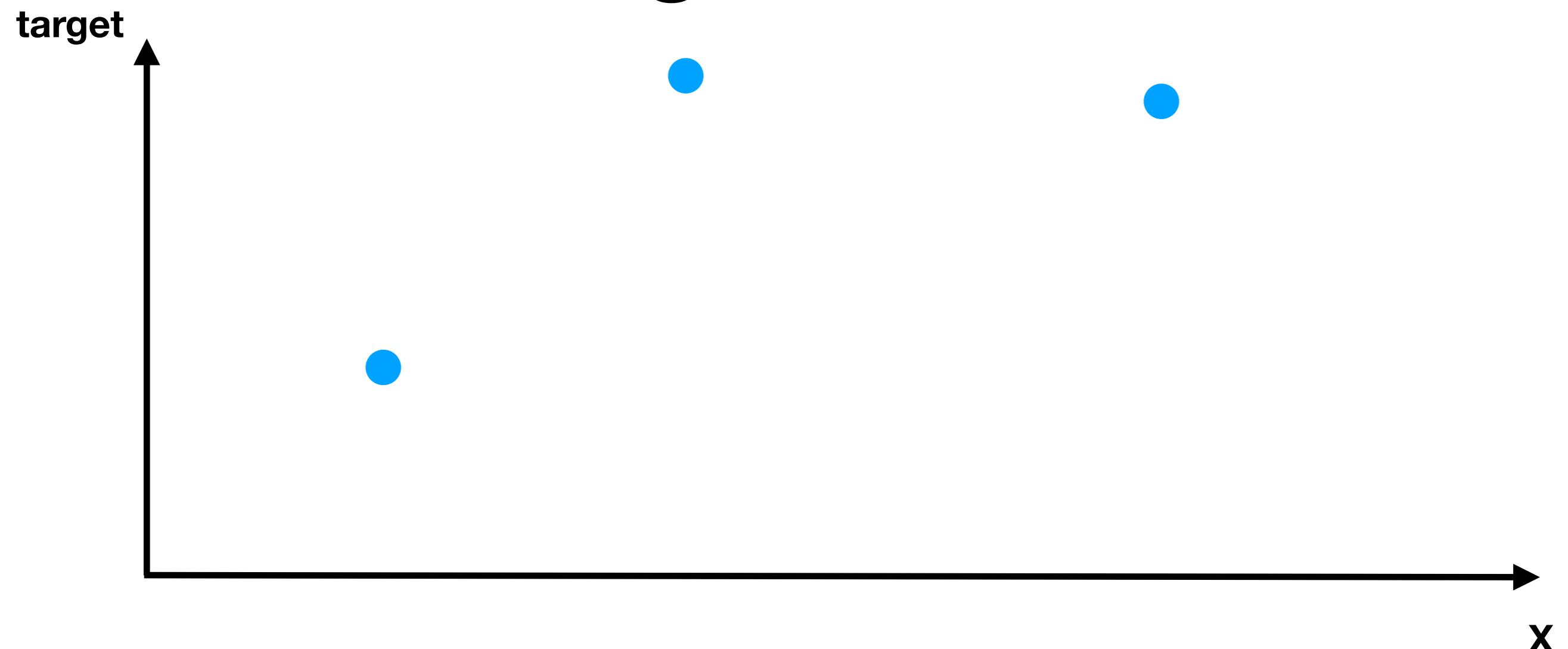
Need find:  $a, b, c, d$

If we find:  $\mathbf{Y1}$  ( $a_1, b_1, c_1, d_1$ ) and  $\mathbf{Y2}$  ( $a_2, b_2, c_2, d_2$ )

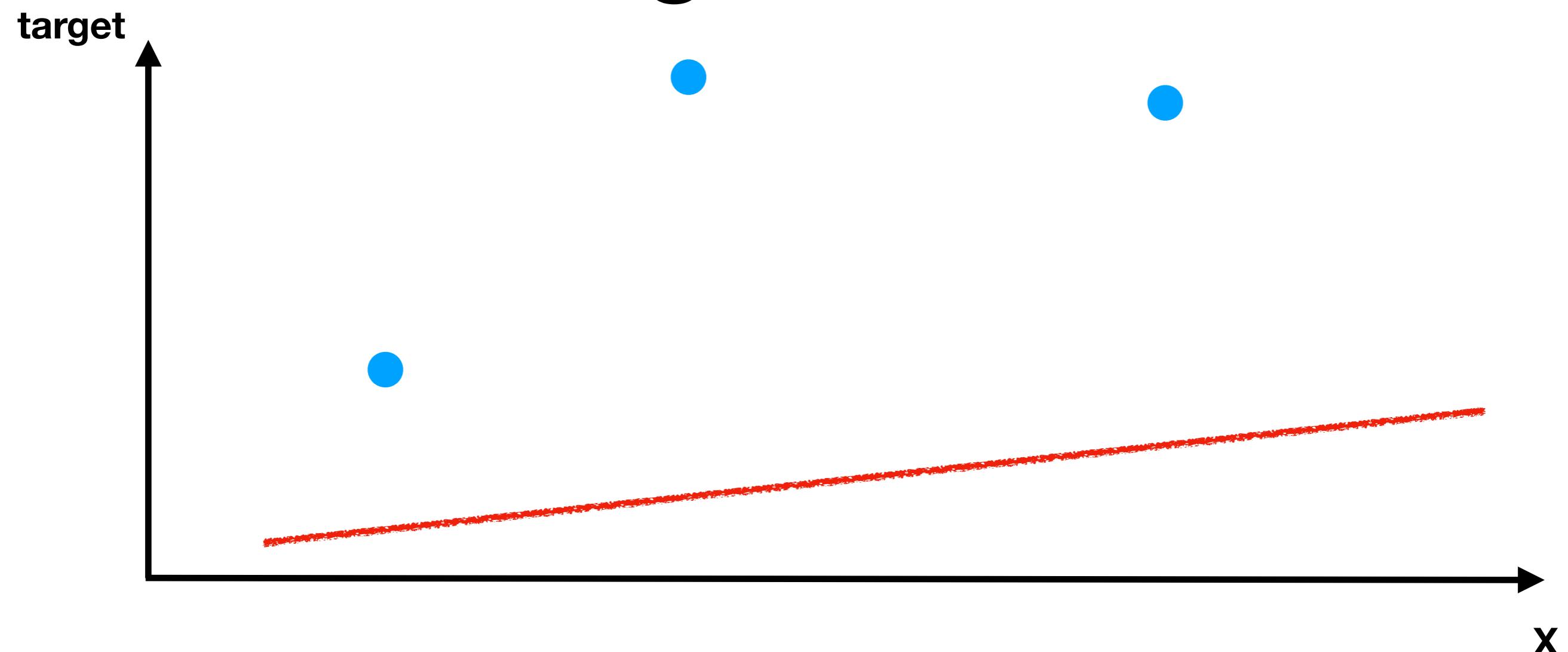
But what's better?  $\mathbf{Y1}$  or  $\mathbf{Y2}$

Need to define **Error**

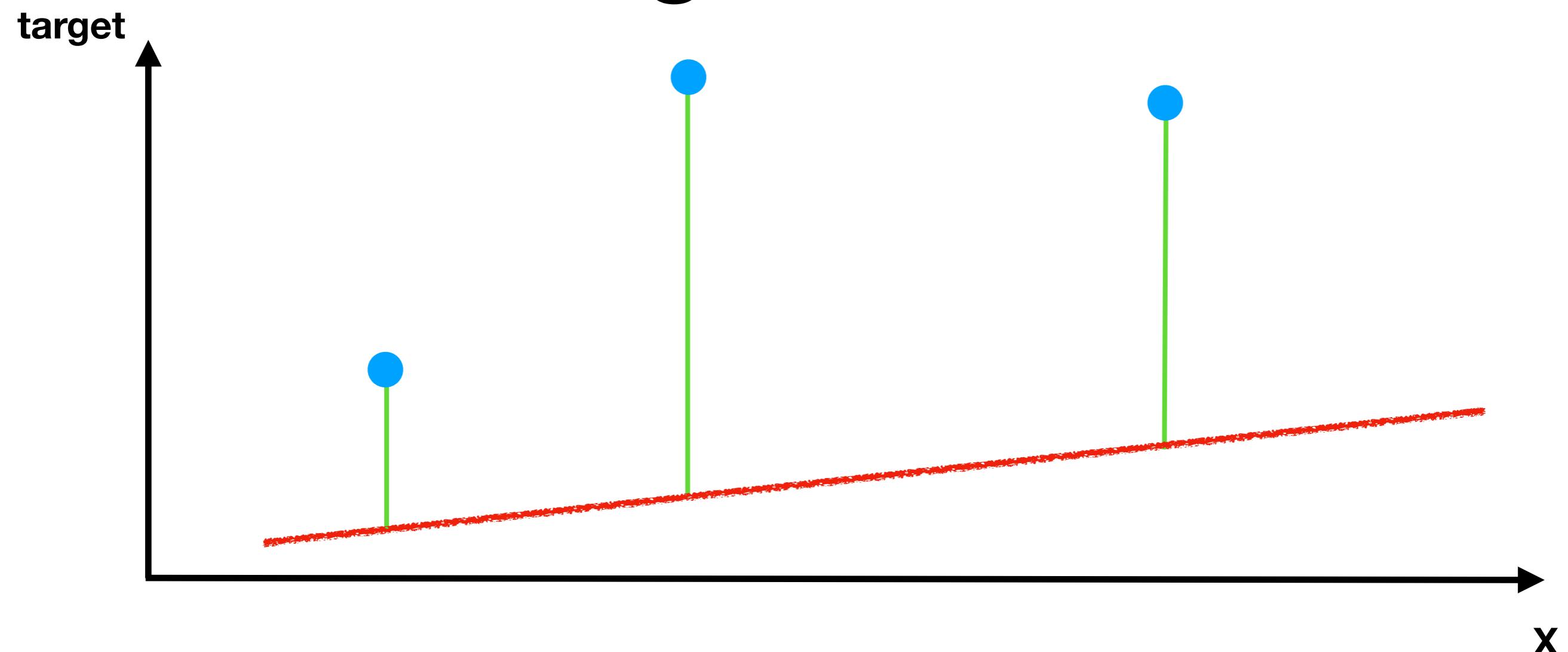
# Good or bad regression



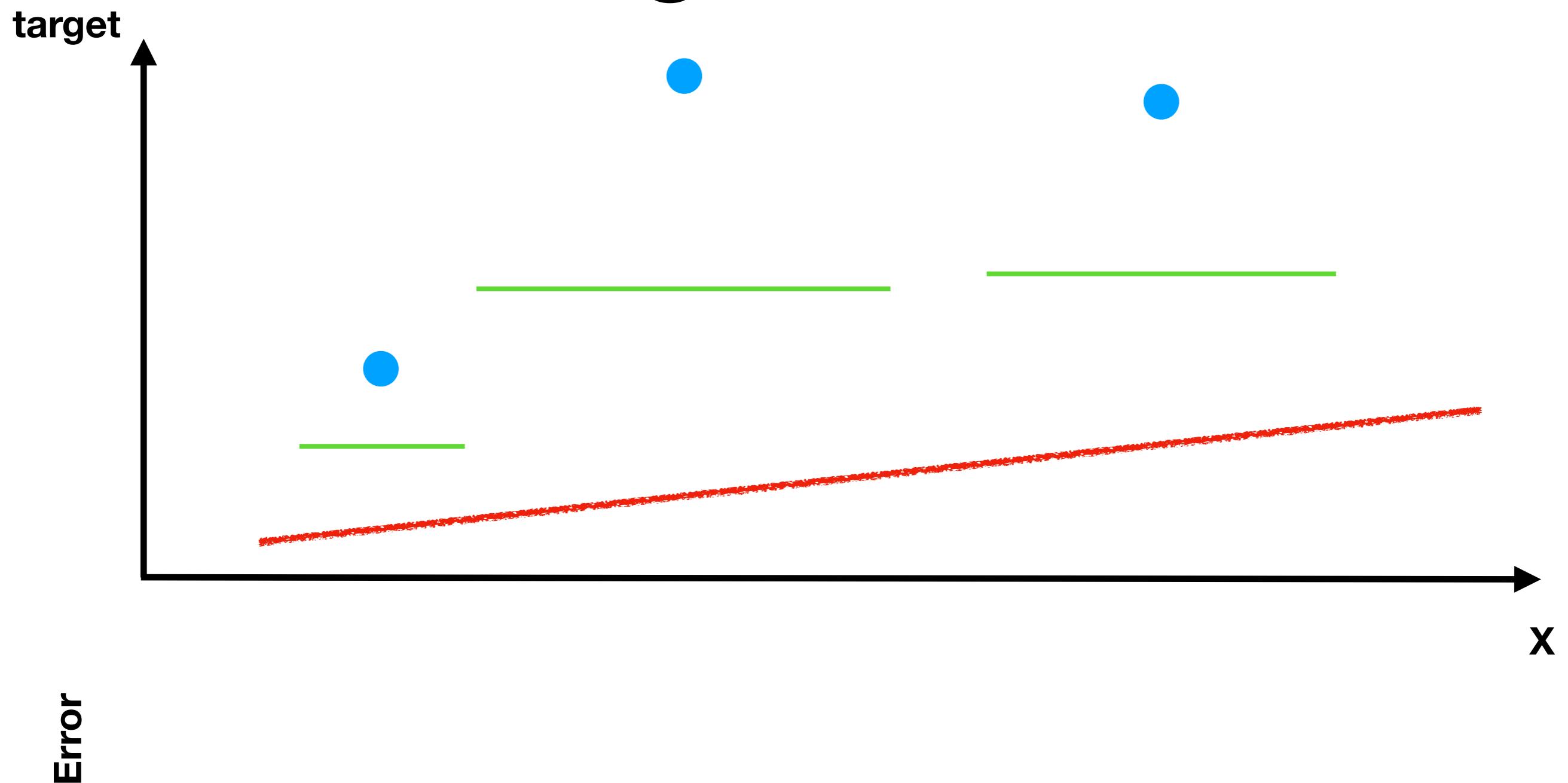
# Good or bad regression



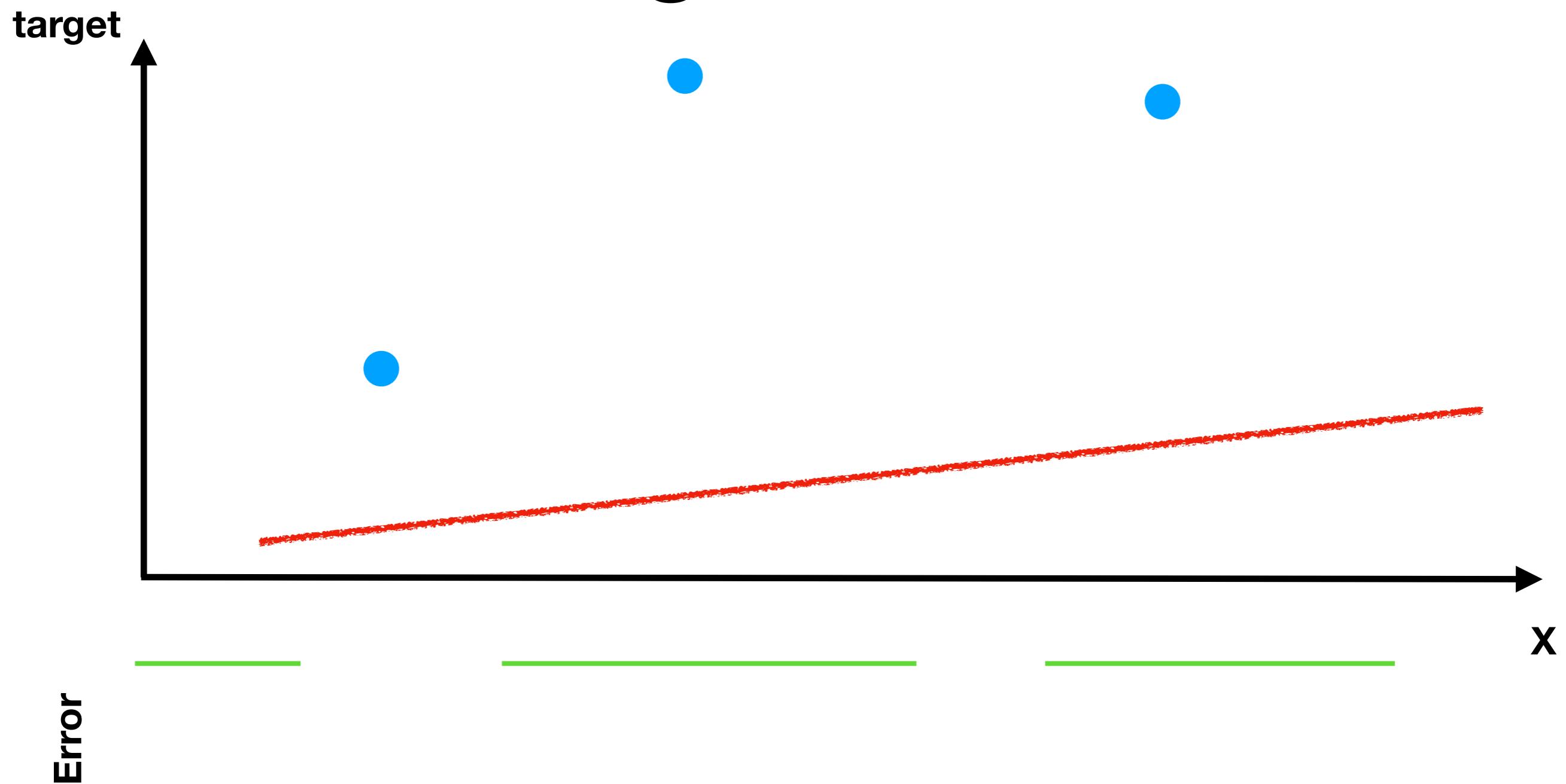
# Good or bad regression



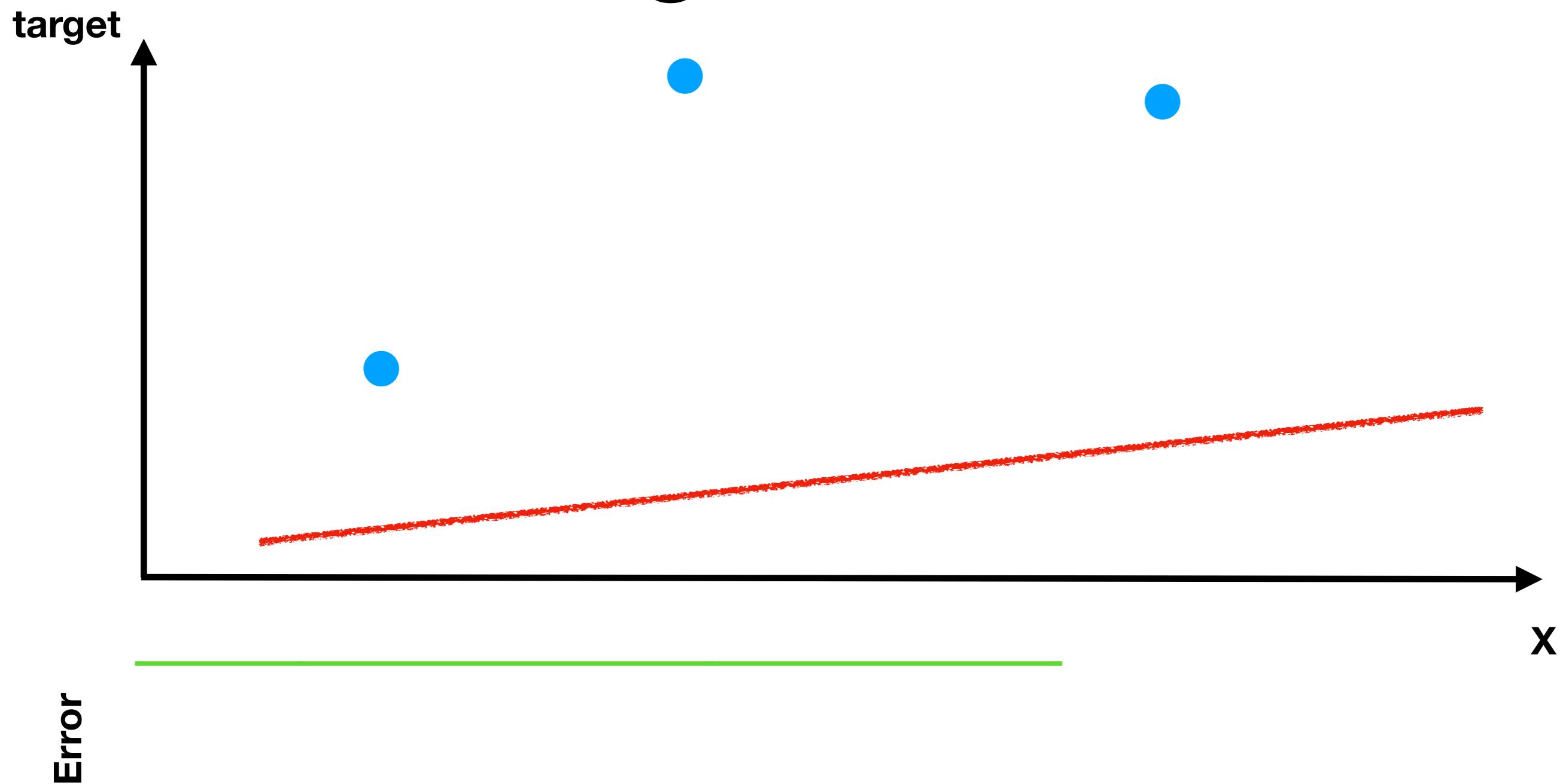
# Good or bad regression



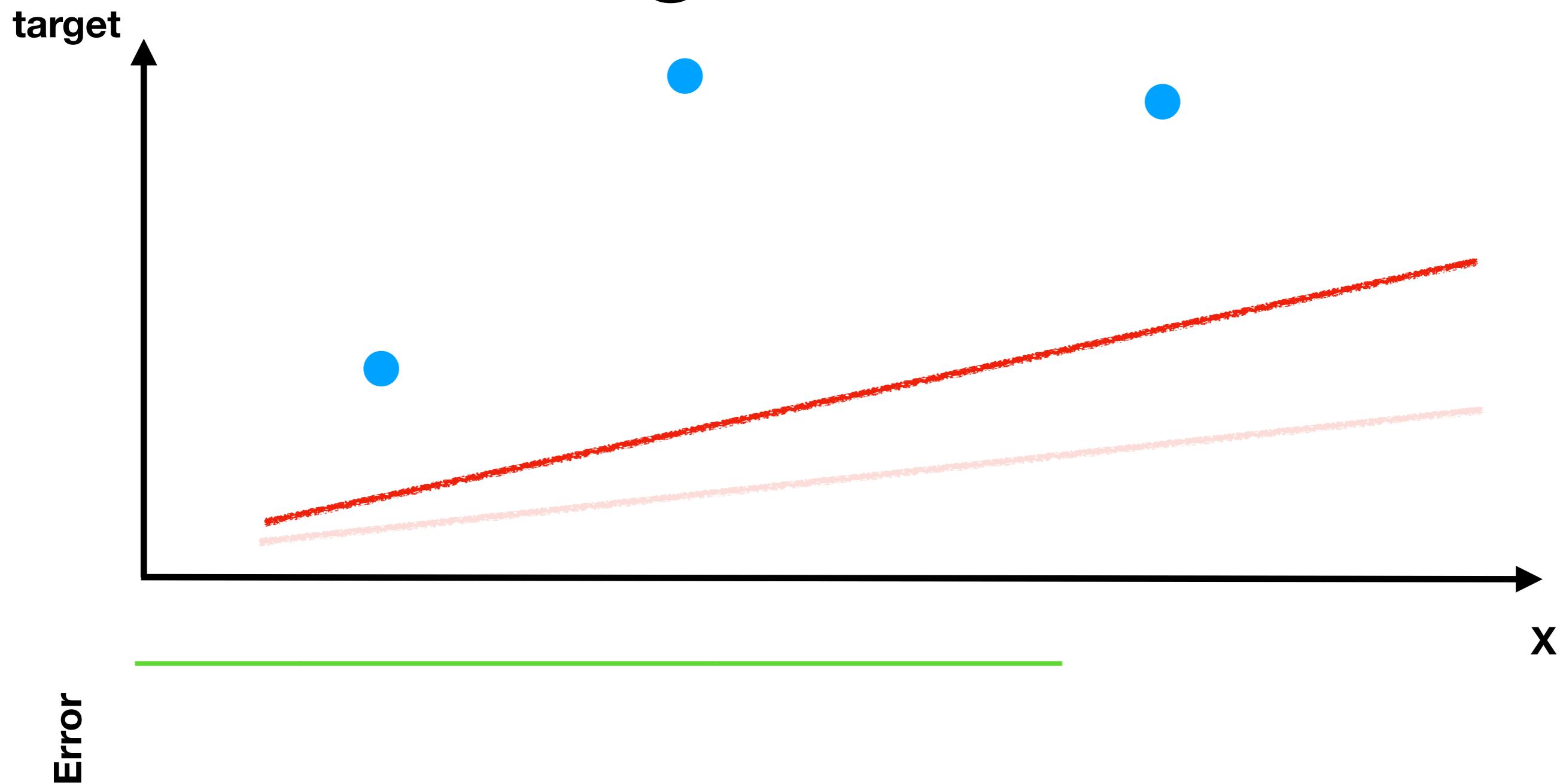
# Good or bad regression



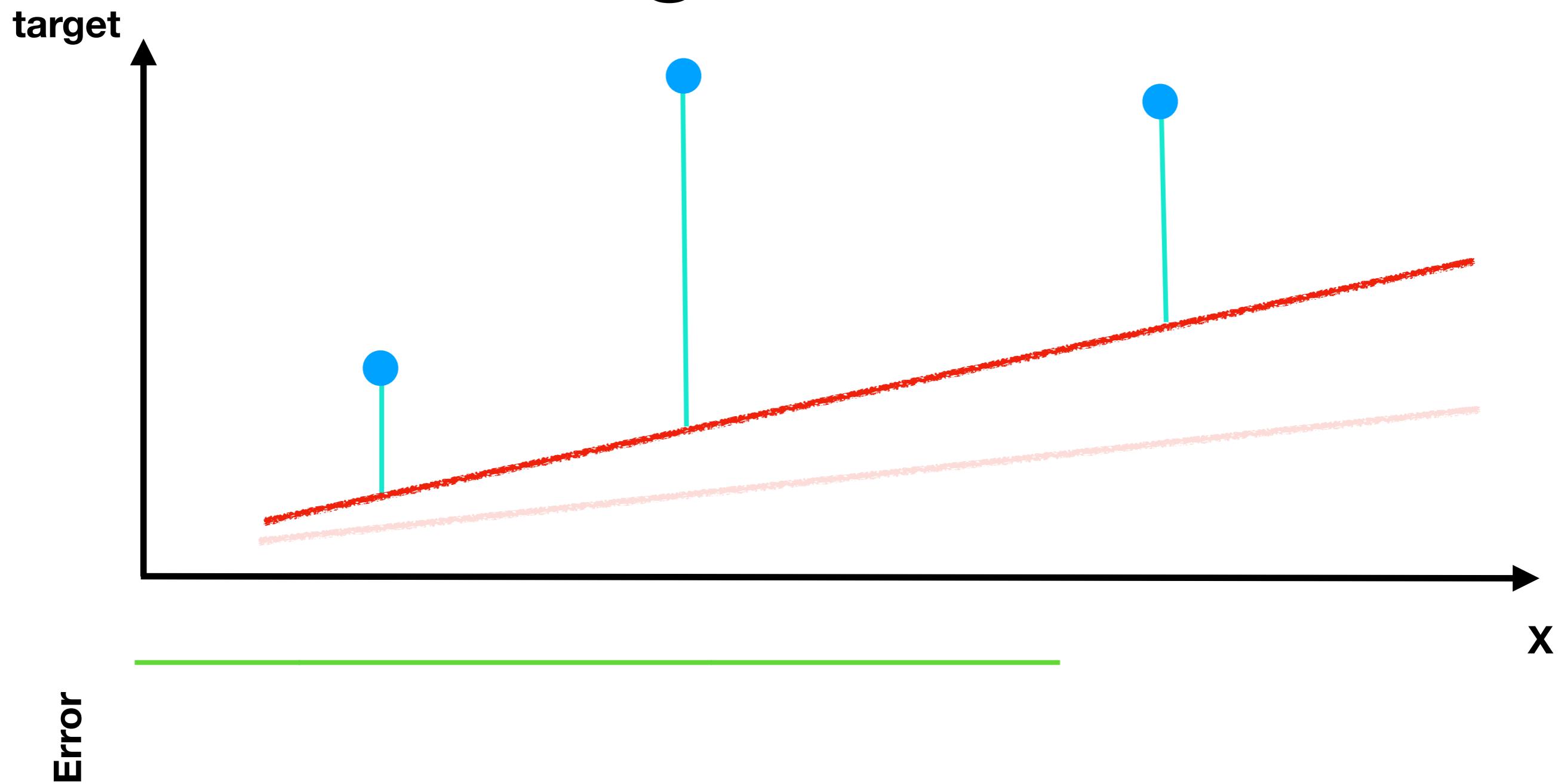
# Good or bad regression



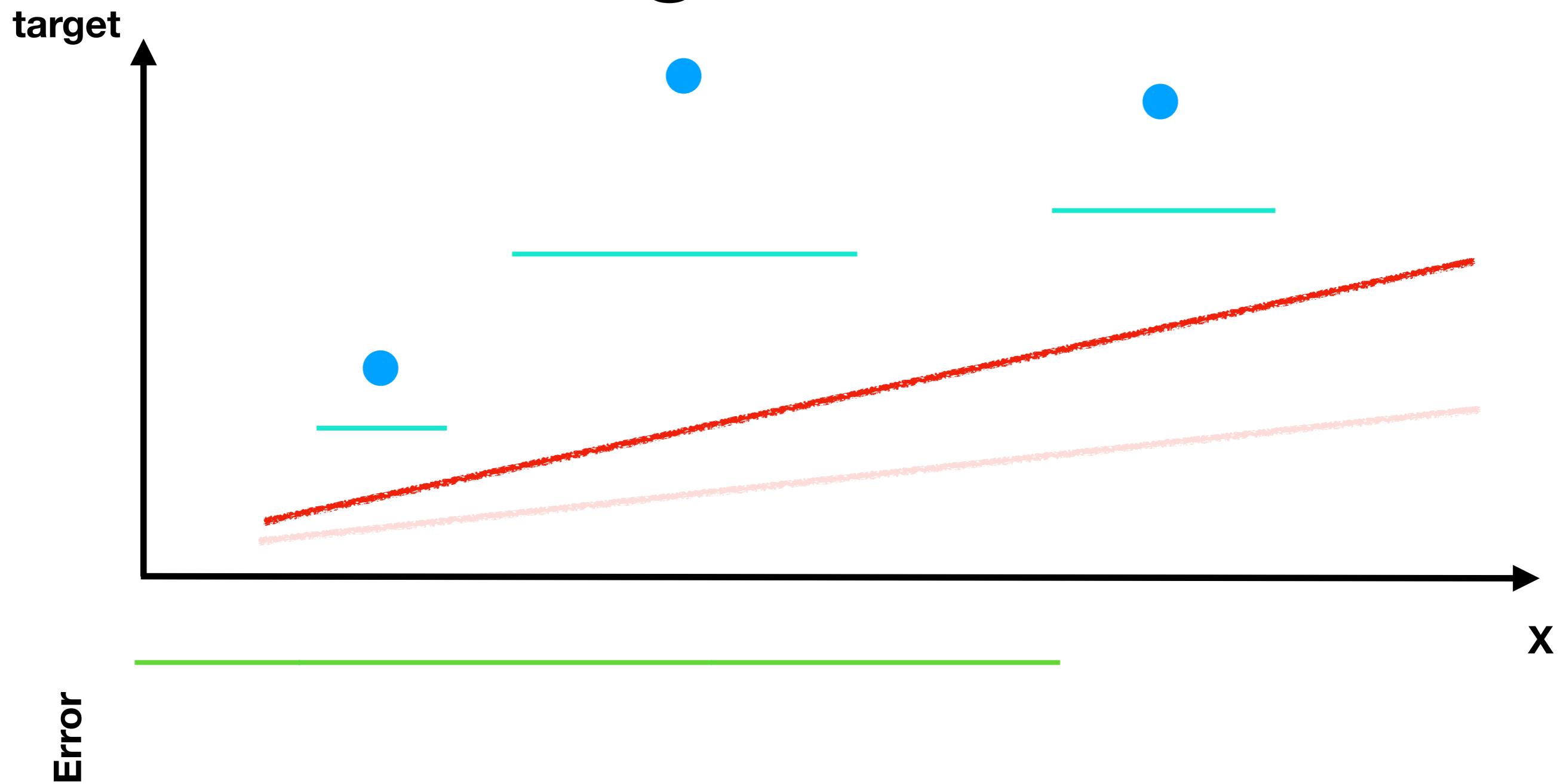
# Good or bad regression



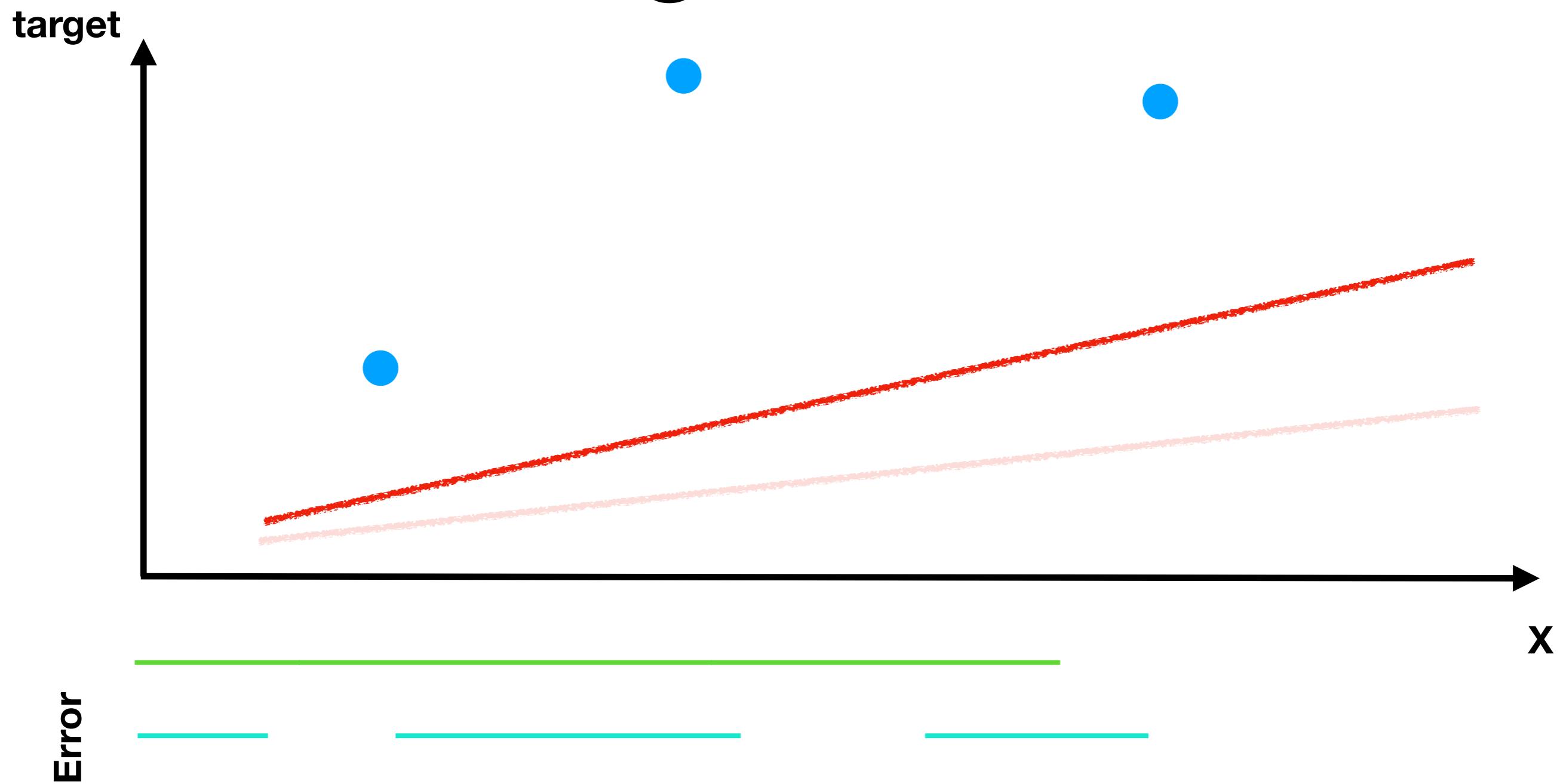
# Good or bad regression



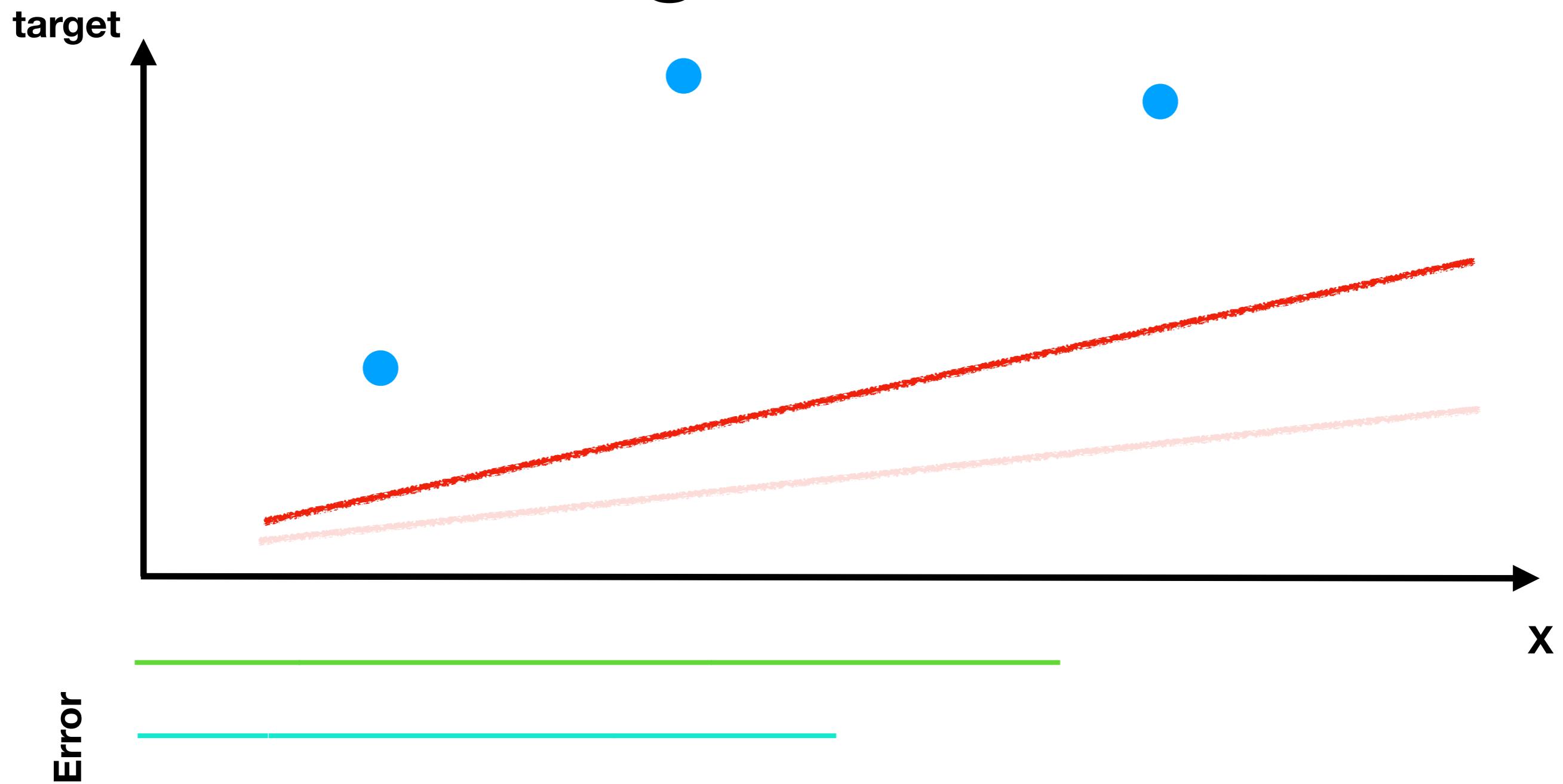
# Good or bad regression



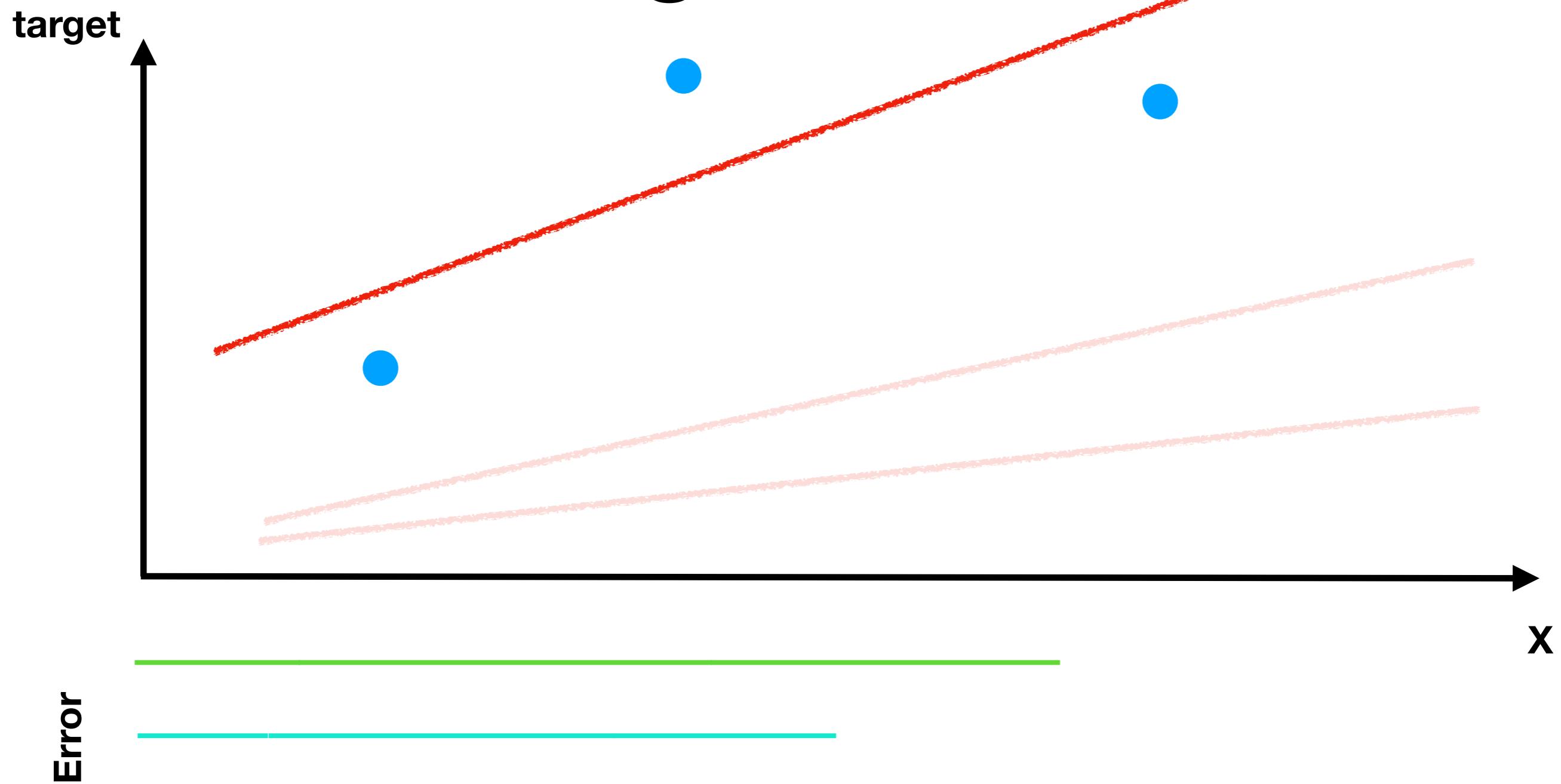
# Good or bad regression



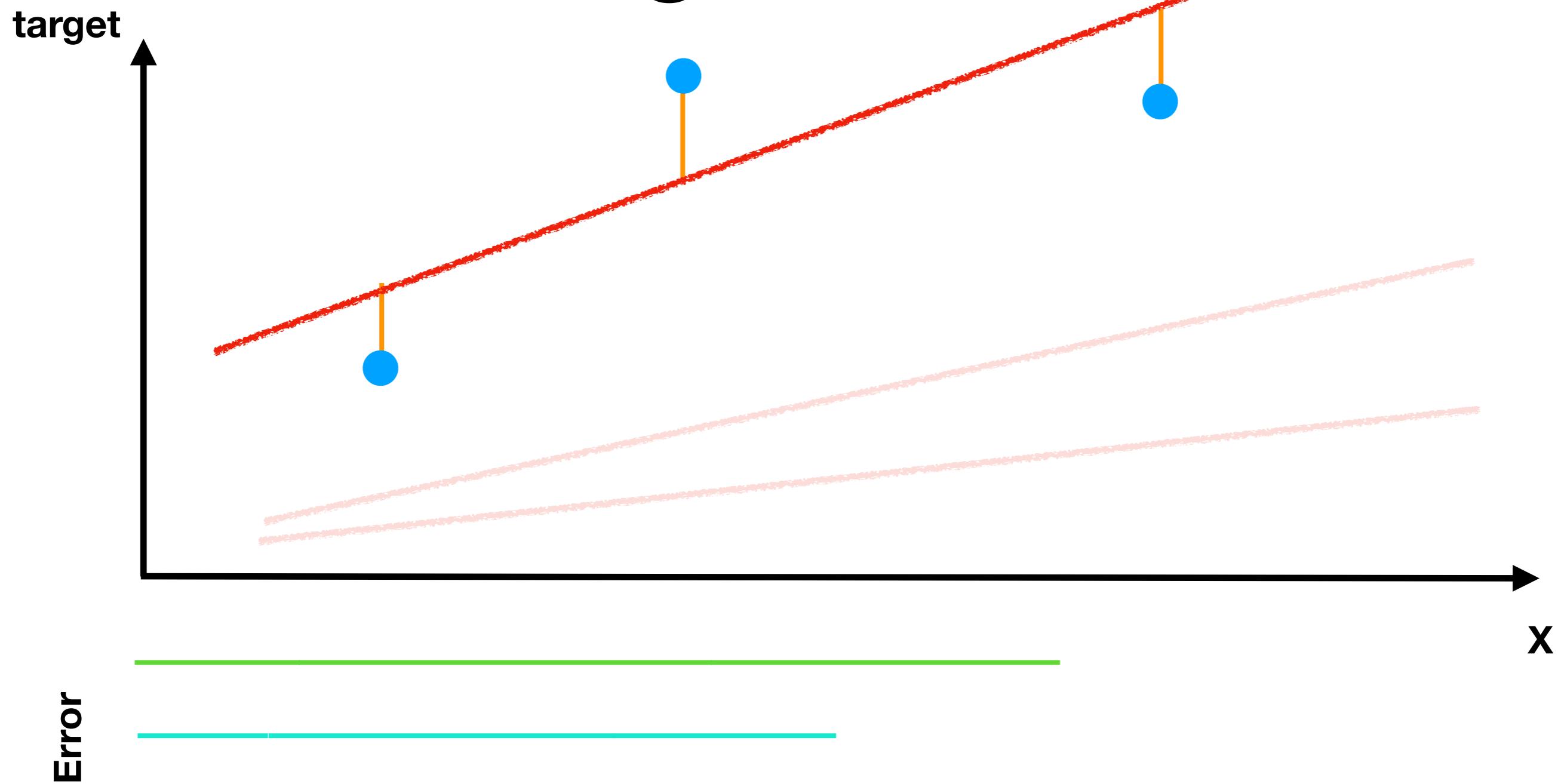
# Good or bad regression



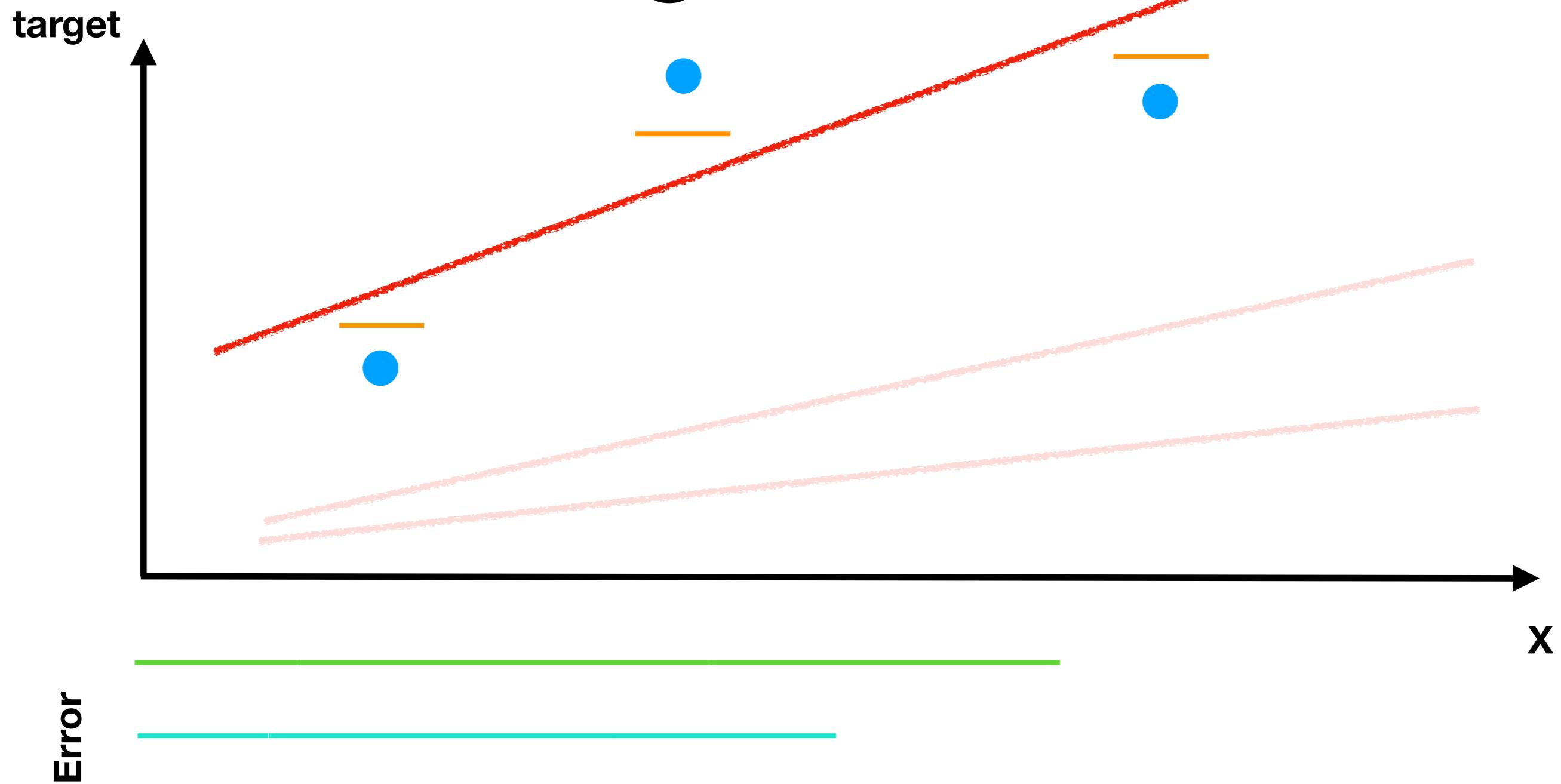
# Good or bad regression



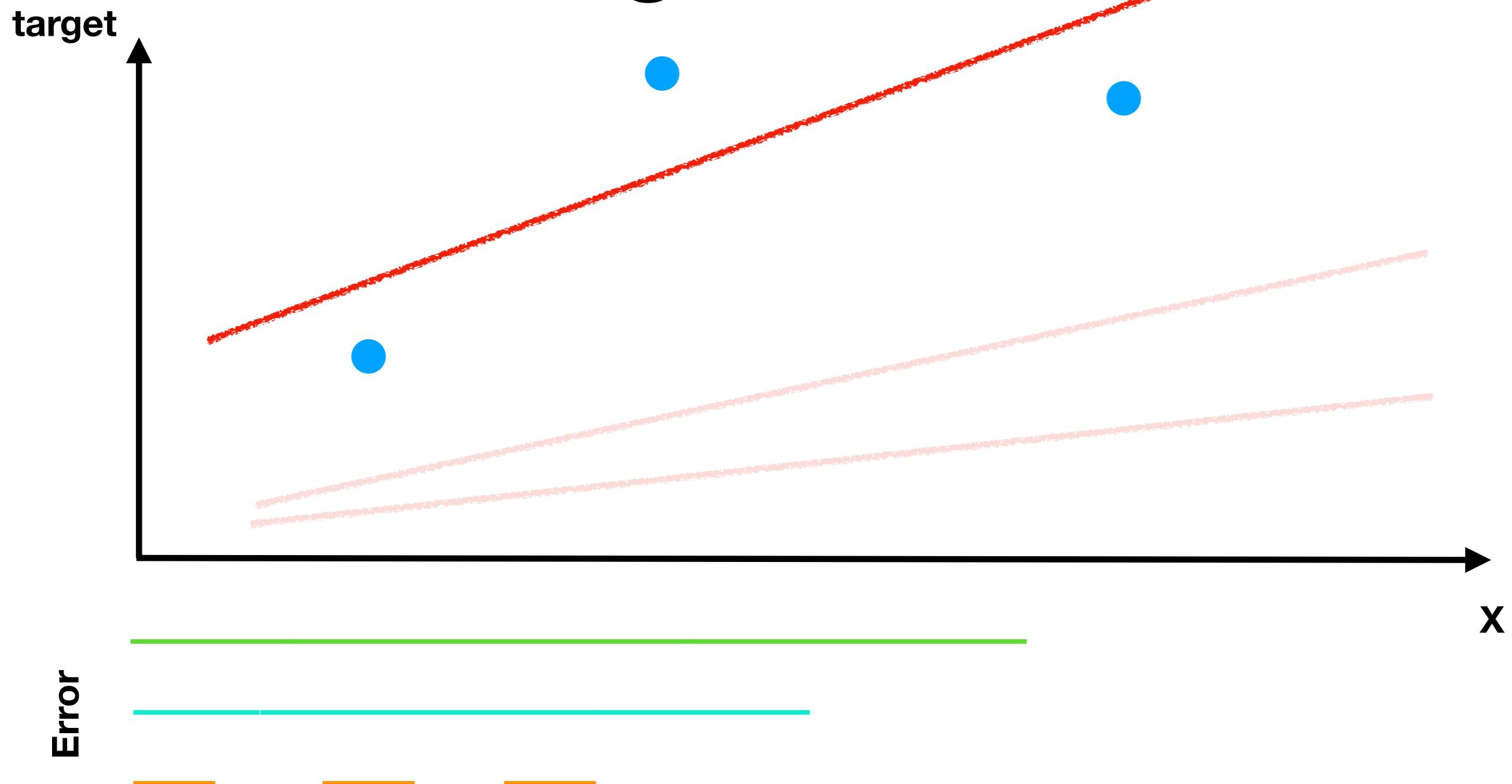
# Good or bad regression



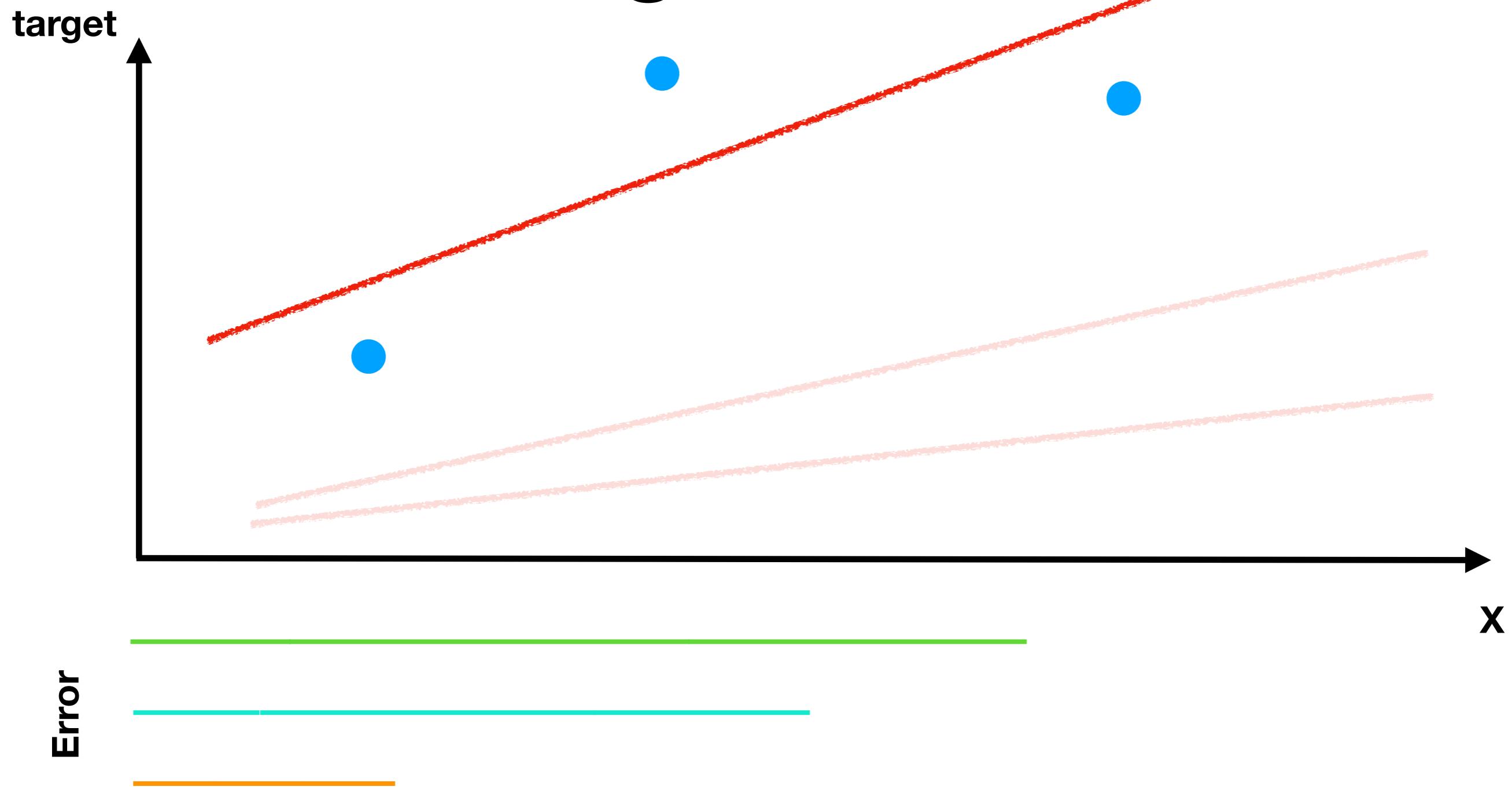
# Good or bad regression



# Good or bad regression



# Good or bad regression



# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	???

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

$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$

$$\hat{Y} = f(X_1, X_2, X_3)$$

$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	???

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

$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$

$$\hat{Y} = f(X_1, X_2, X_3)$$

$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

**Machine Learning Task:**

find:  $a, b, c, d$

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	???

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

$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$

$$\hat{Y} = f(X_1, X_2, X_3)$$

$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

## Machine Learning Task:

find:  $a, b, c, d$

and minimization of error

# Regression metrics

# Regression metrics

- MAE

# Regression metrics

- MAE

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

# Regression metrics

- MAE

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

- MSE

# Regression metrics

- MAE

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

- MSE

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

# Regression metrics

- MAE

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

- MSE

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

- RMSE

# Regression metrics

- MAE

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

- MSE

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

- RMSE

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

# Regression metrics

- MAE

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

- MSE

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

- RMSE

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- RMSLE

# Regression metrics

- MAE

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

- MSE

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

- RMSE

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- RMSLE

$$\text{RMSLE} = \sqrt{\frac{1}{n} \sum_{j=1}^n (\log(y_j + 1) - \log(\hat{y}_j + 1))^2}$$

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	???

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

$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$

$$\hat{Y} = f(X_1, X_2, X_3)$$

$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

Ok! We find best **parameters**:  $a_{best}, b_{best}, c_{best}, d_{best}$

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	???

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

$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$

$$\hat{Y} = f(X_1, X_2, X_3)$$

$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

Ok! We find best **parameters**:  $a_{best}, b_{best}, c_{best}, d_{best}$

$$\hat{Y} = a_{best}^*X_1 + b_{best}^*X_2 + c_{best}^*X_3 + d_{best}$$

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	???

$$\hat{Y} = f(X)$$
$$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$$
$$\hat{Y} = f(X_1, X_2, X_3)$$
$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

Ok! We find best parameters:  $a_{best}, b_{best}, c_{best}, d_{best}$

$$\hat{Y} = a_{best}^*X_1 + b_{best}^*X_2 + c_{best}^*X_3 + d_{best}$$

$$\hat{y} = a_{best}^*55 + b_{best}^*1 + c_{best}^*25 + d_{best}$$

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	???

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

$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$

$$\hat{Y} = f(X_1, X_2, X_3)$$

$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

Ok! We find best **parameters**:  $a_{best}, b_{best}, c_{best}, d_{best}$

$$\hat{Y} = a_{best}^*X_1 + b_{best}^*X_2 + c_{best}^*X_3 + d_{best}$$

$$\hat{y} = a_{best}^*55 + b_{best}^*1 + c_{best}^*25 + d_{best} = 92$$

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	???

$$\hat{Y} = f(X)$$
$$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$$
$$\hat{Y} = f(X_1, X_2, X_3)$$
$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

Ok! We find best **parameters**:  $a_{best}, b_{best}, c_{best}, d_{best}$

$$\hat{Y} = a_{best}^*X_1 + b_{best}^*X_2 + c_{best}^*X_3 + d_{best}$$

$$\hat{y} = a_{best}^*55 + b_{best}^*1 + c_{best}^*25 + d_{best} = 92$$

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	92

$$\hat{Y} = f(X)$$
$$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$$
$$\hat{Y} = f(X_1, X_2, X_3)$$
$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

Ok! We find best **parameters**:  $a_{best}, b_{best}, c_{best}, d_{best}$

$$\hat{Y} = a_{best}^*X_1 + b_{best}^*X_2 + c_{best}^*X_3 + d_{best}$$

$$\hat{y} = a_{best}^*55 + b_{best}^*1 + c_{best}^*25 + d_{best} = 92$$

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	92

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

$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$

$$\hat{Y} = f(X_1, X_2, X_3)$$

$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

Ok! We find best **parameters**:  $a_{best}, b_{best}, c_{best}, d_{best}$

$$\hat{Y} = a_{best}^*X_1 + b_{best}^*X_2 + c_{best}^*X_3 + d_{best}$$

$$\hat{y} = a_{best}^*55 + b_{best}^*1 + c_{best}^*25 + d_{best} = 92$$

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	92

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

$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$

$$\hat{Y} = f(X_1, X_2, X_3)$$

$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

## Linear regression

# Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	Price (thous. \$)
70	2	40	100
130	3	15	160
40	1	5	83
55	1	25	92

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

$\hat{Y} = f(\text{total area, number of bedrooms, age of the house})$

$$\hat{Y} = f(X_1, X_2, X_3)$$

$$\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$$

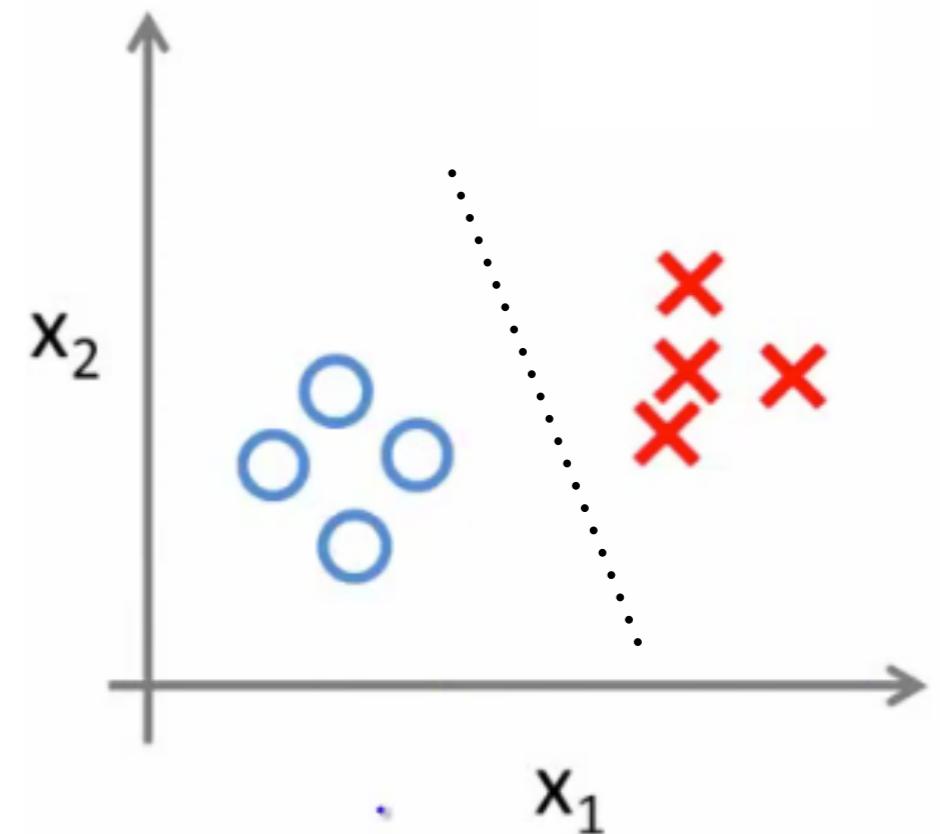
All values of features are real numbers

# Classification

- $\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$  - *regression example*

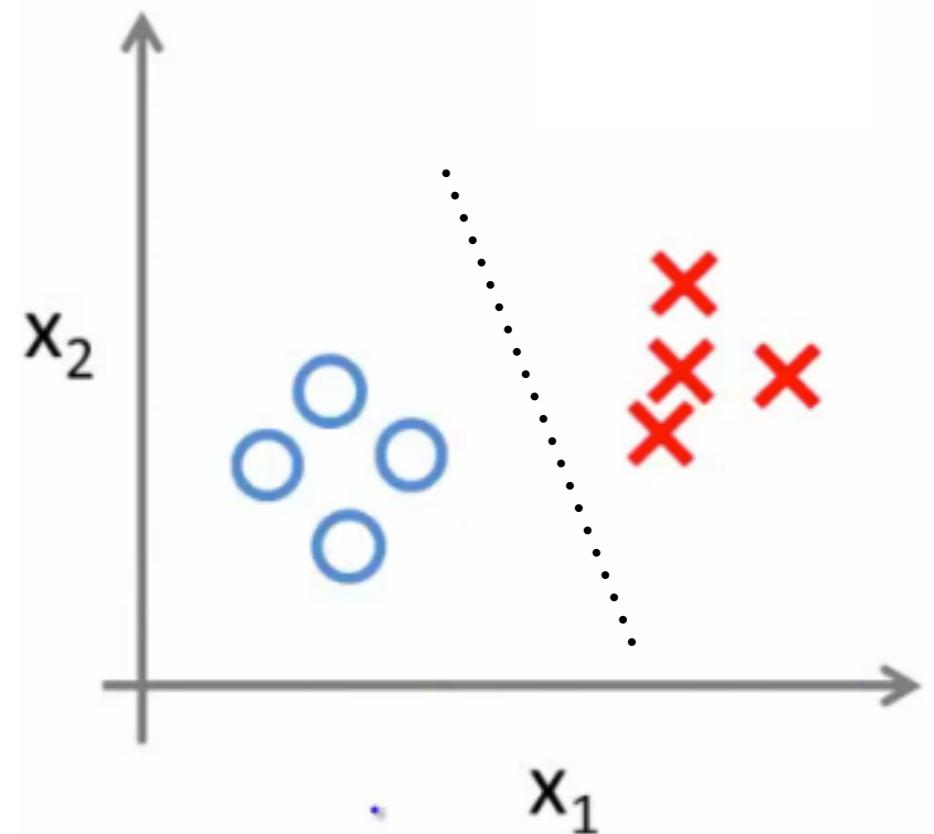
# Classification

- $\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$  - regression example



# Classification

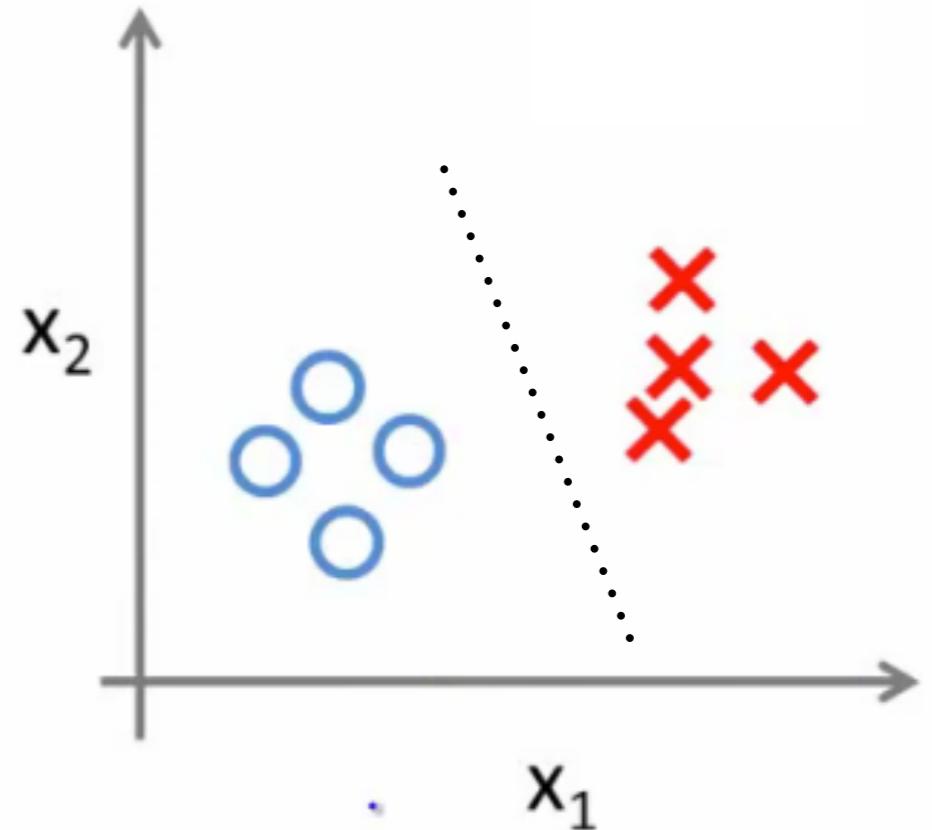
- $\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$  - regression example
- $\hat{Y} = f(X)$



# Classification

- $\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$  - regression example
- $\hat{Y} = f(X)$

$$\hat{Y} = a^*X_1 + b^*X_2 + c$$



# Classification

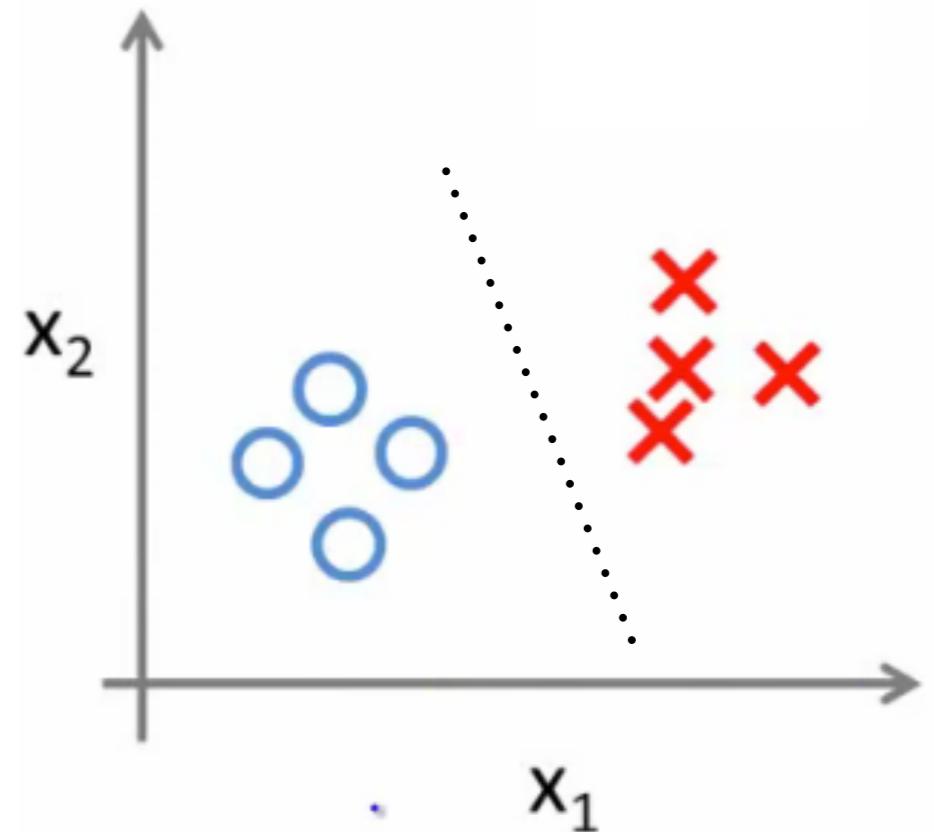
- $\hat{Y} = a^*X_1 + b^*X_2 + c^*X_3 + d$  - regression example

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

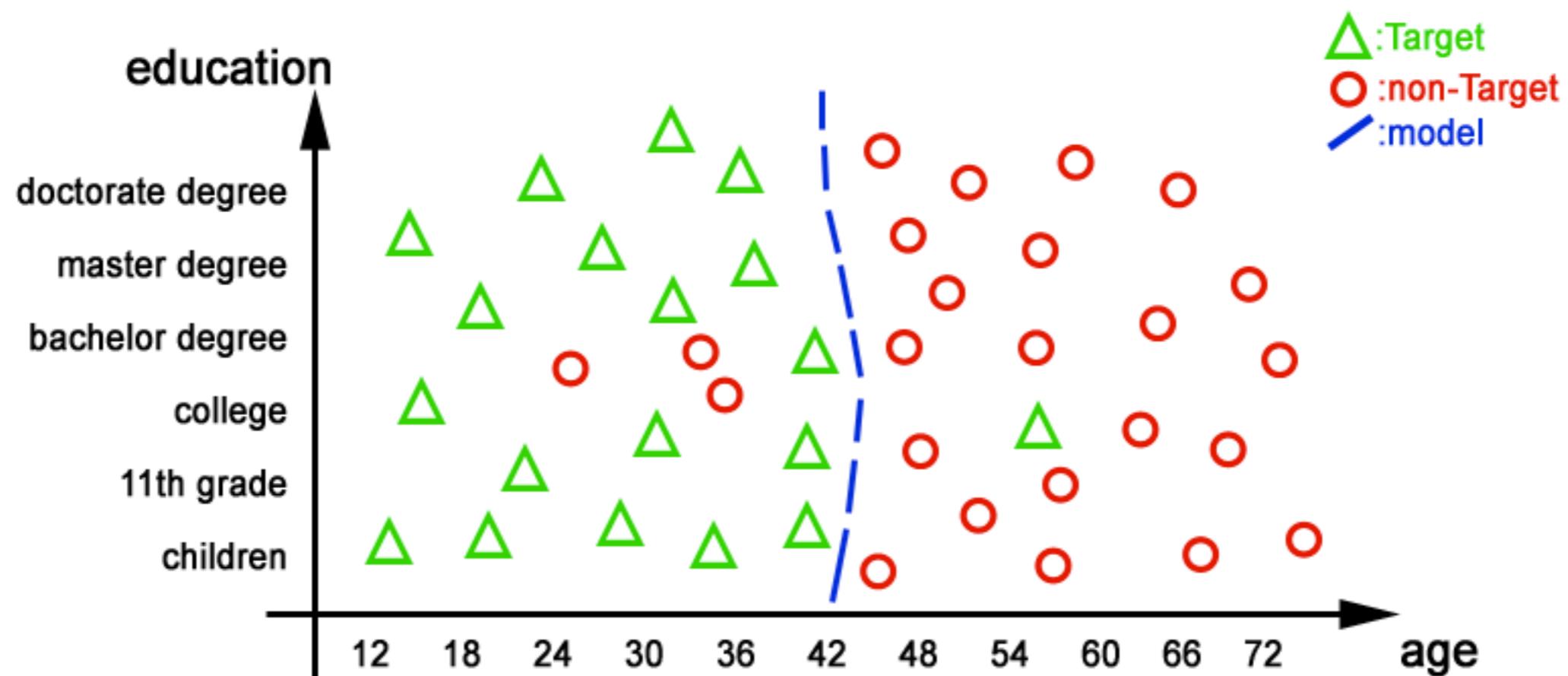
$$\hat{Y} = a^*X_1 + b^*X_2 + c$$

$$\hat{Y} = f(X) > 0 \quad \text{X}$$

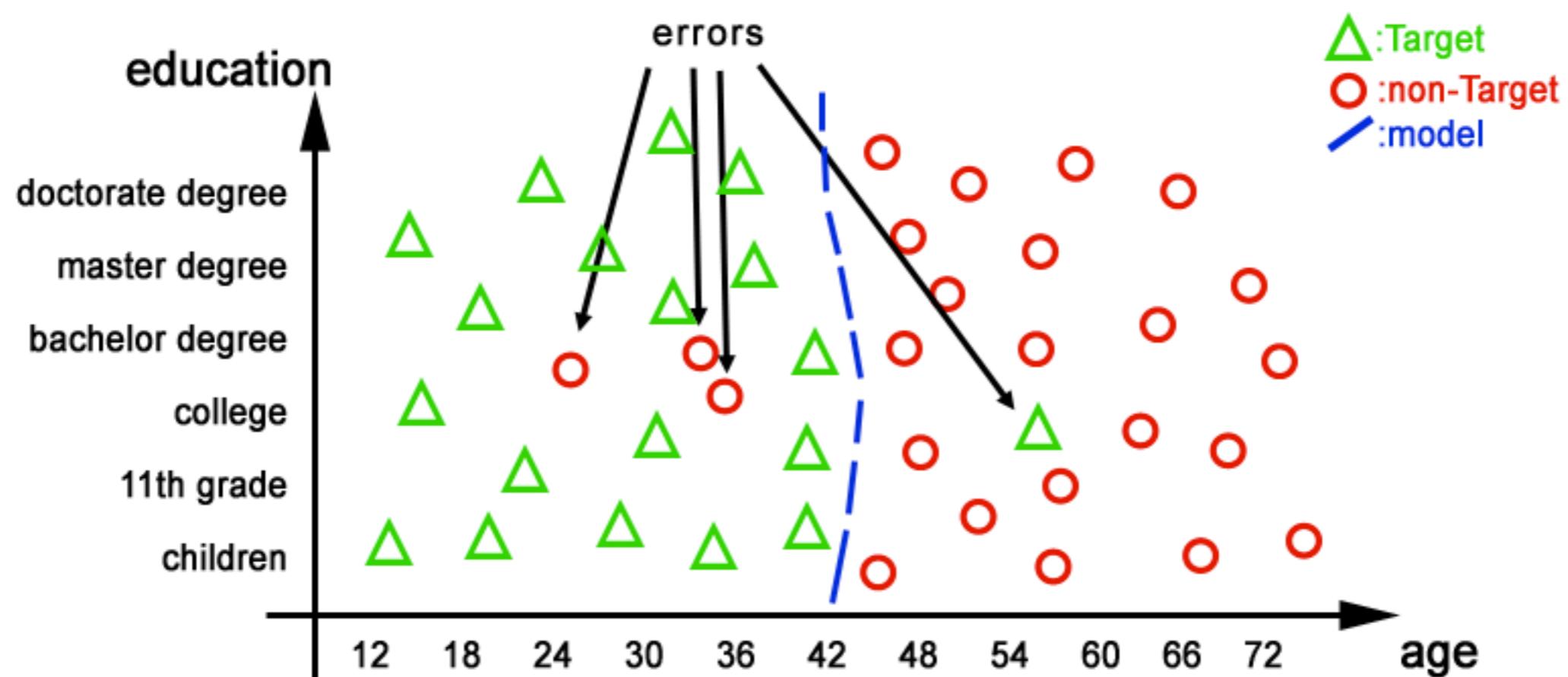
$$\hat{Y} = f(X) < 0 \quad \text{O}$$



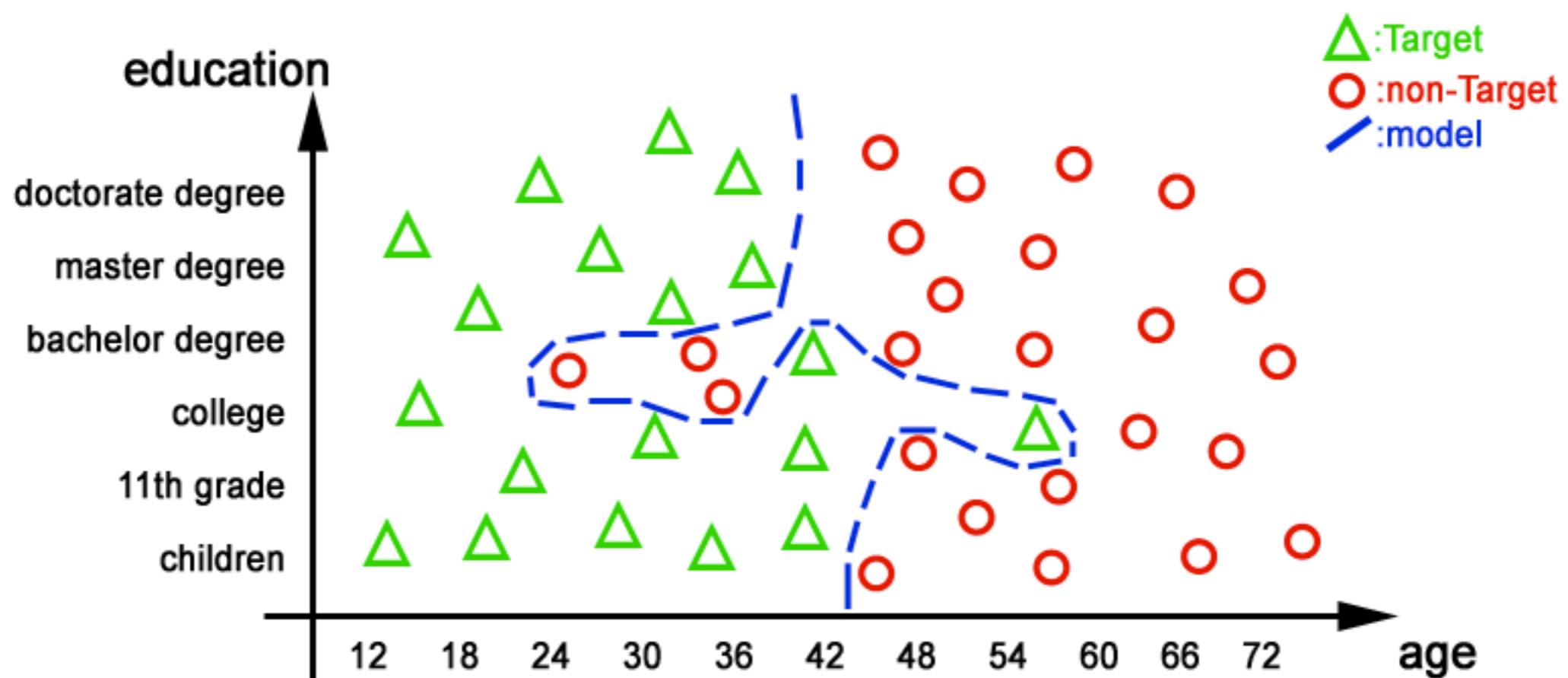
# Model with errors



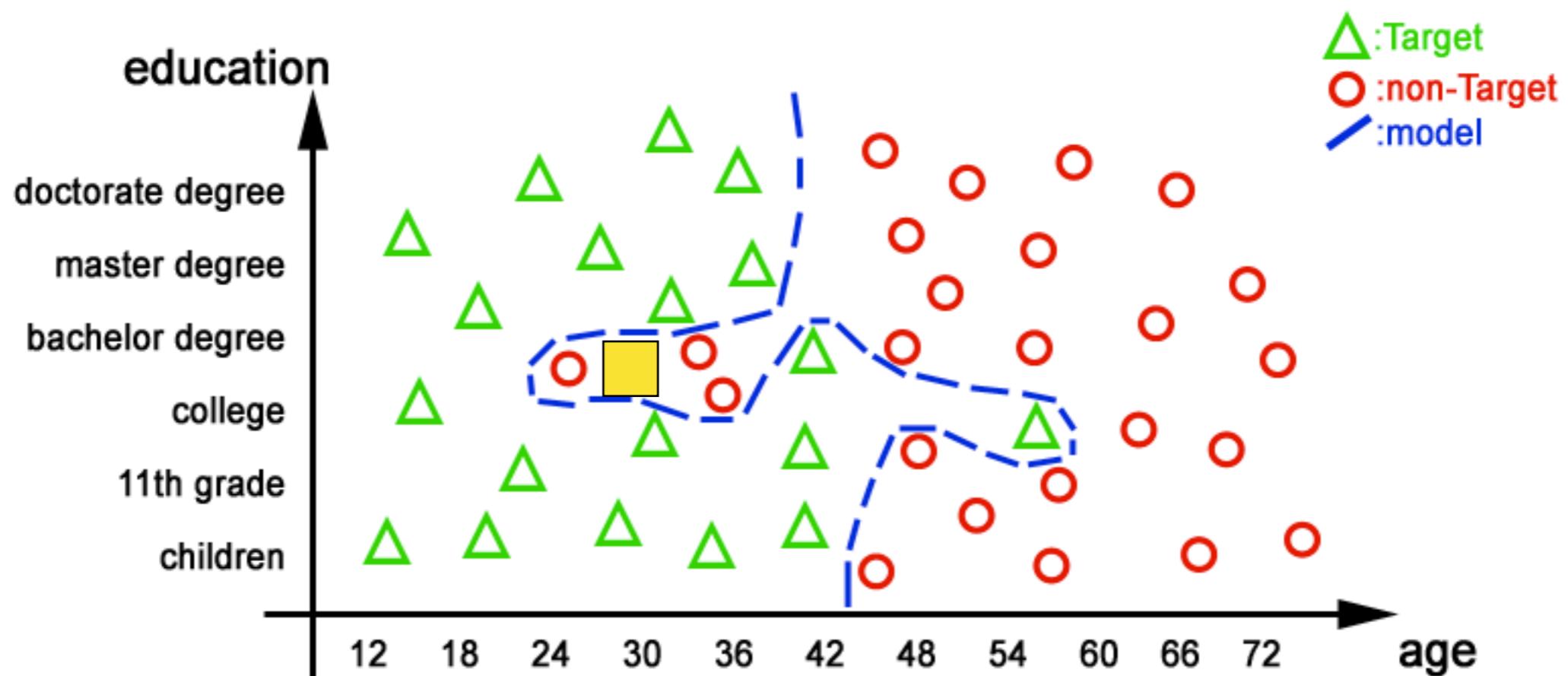
# Model with errors



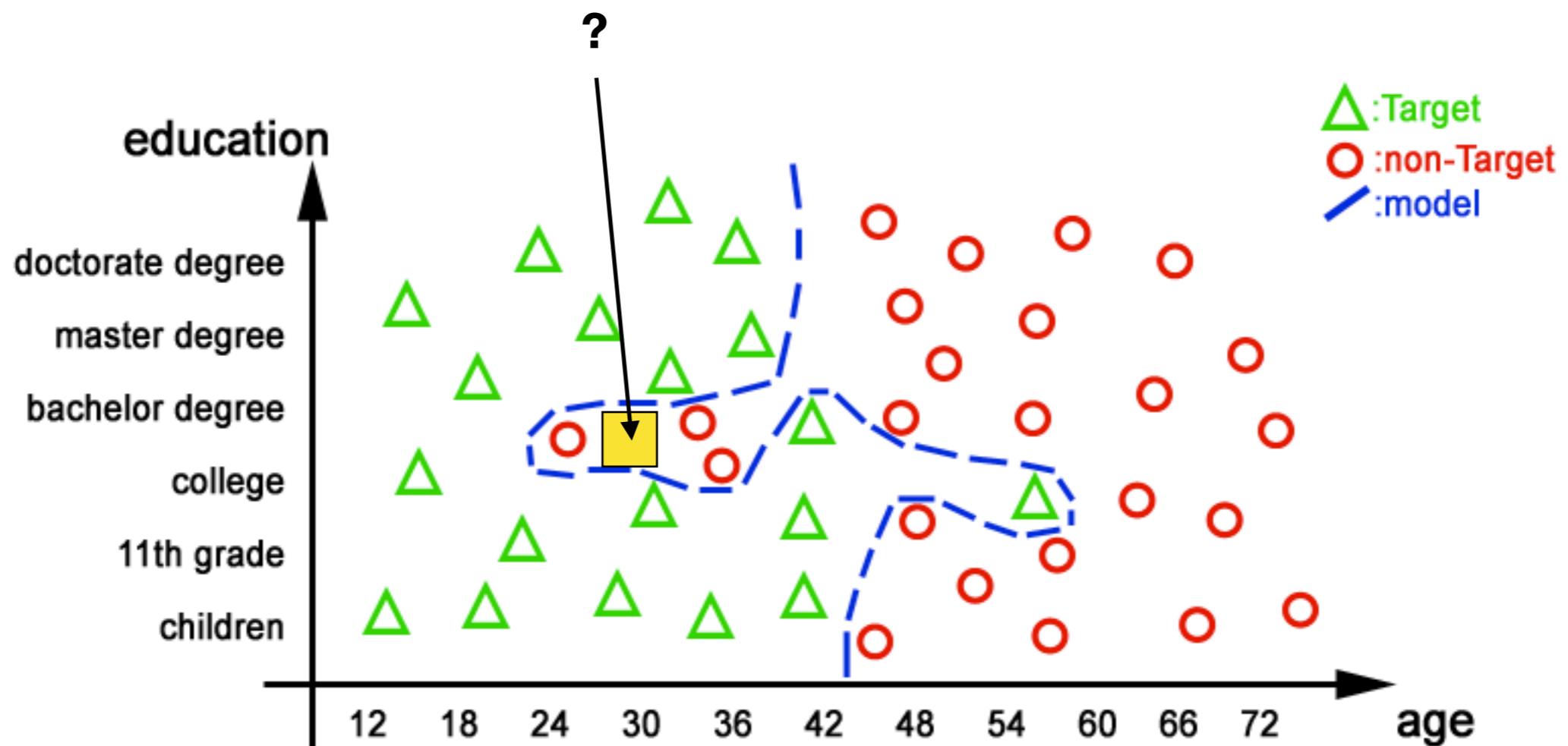
# Overfitting



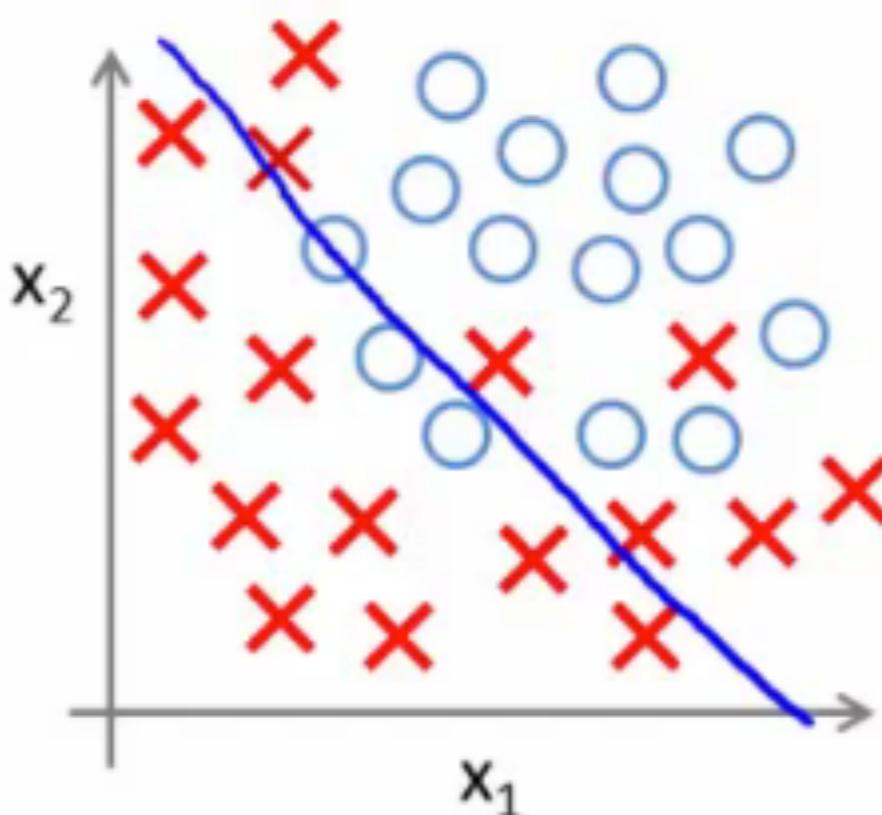
# Overfitting



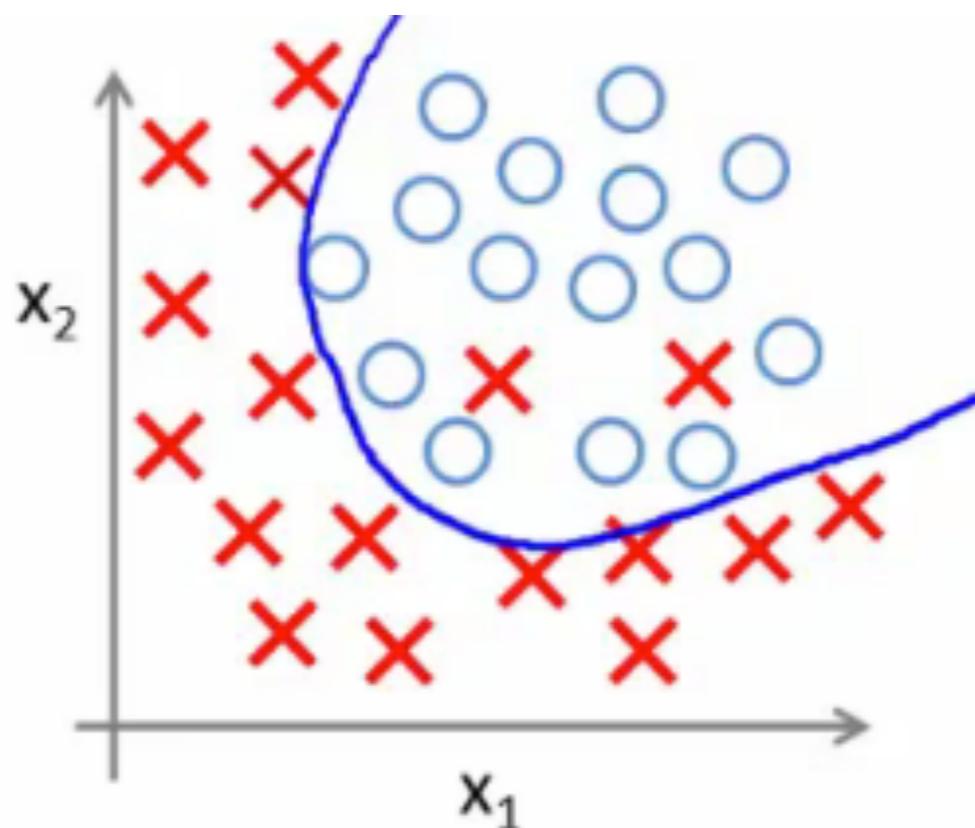
# Overfitting



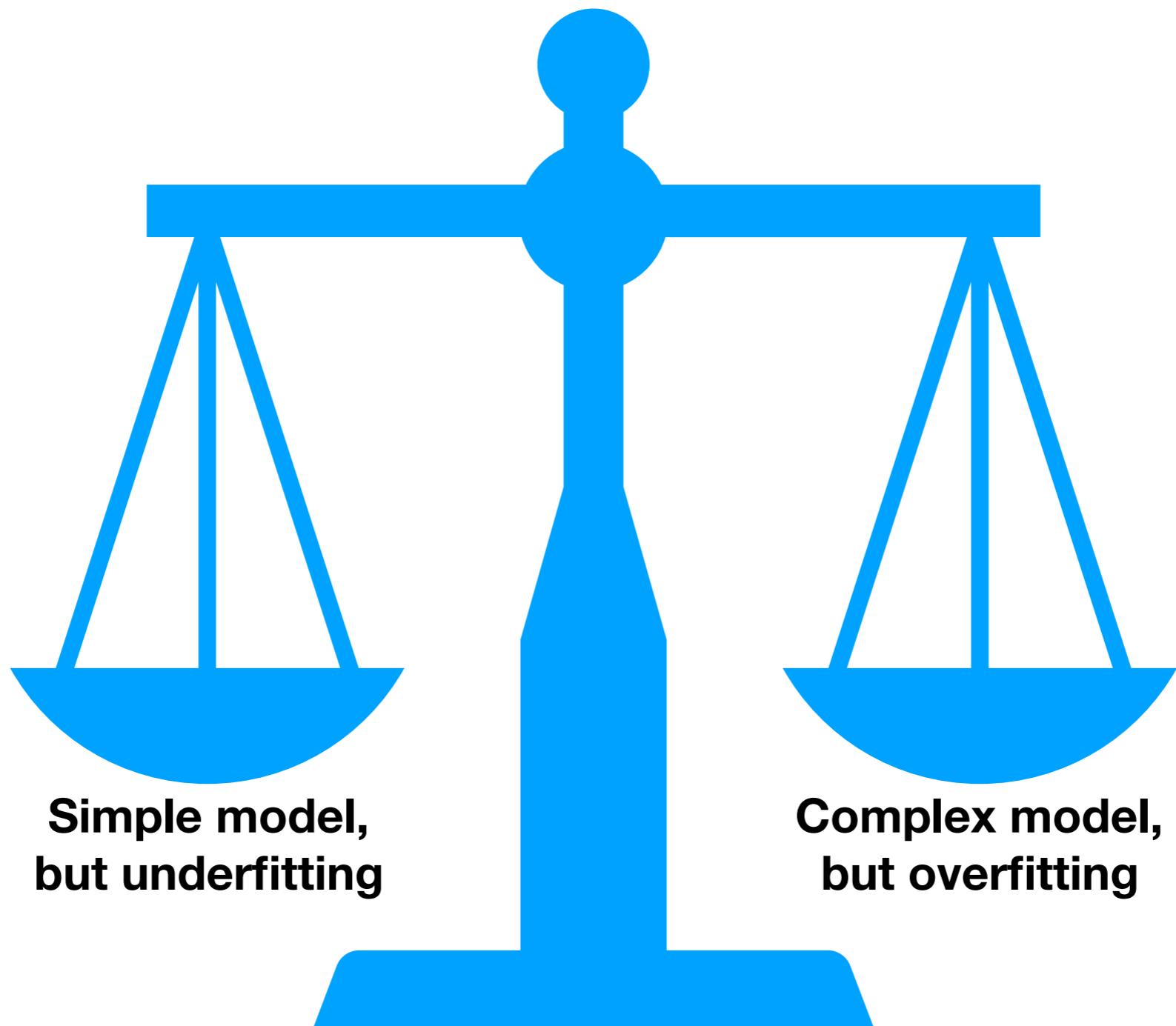
# Underfitting



# Underfitting



# Best model



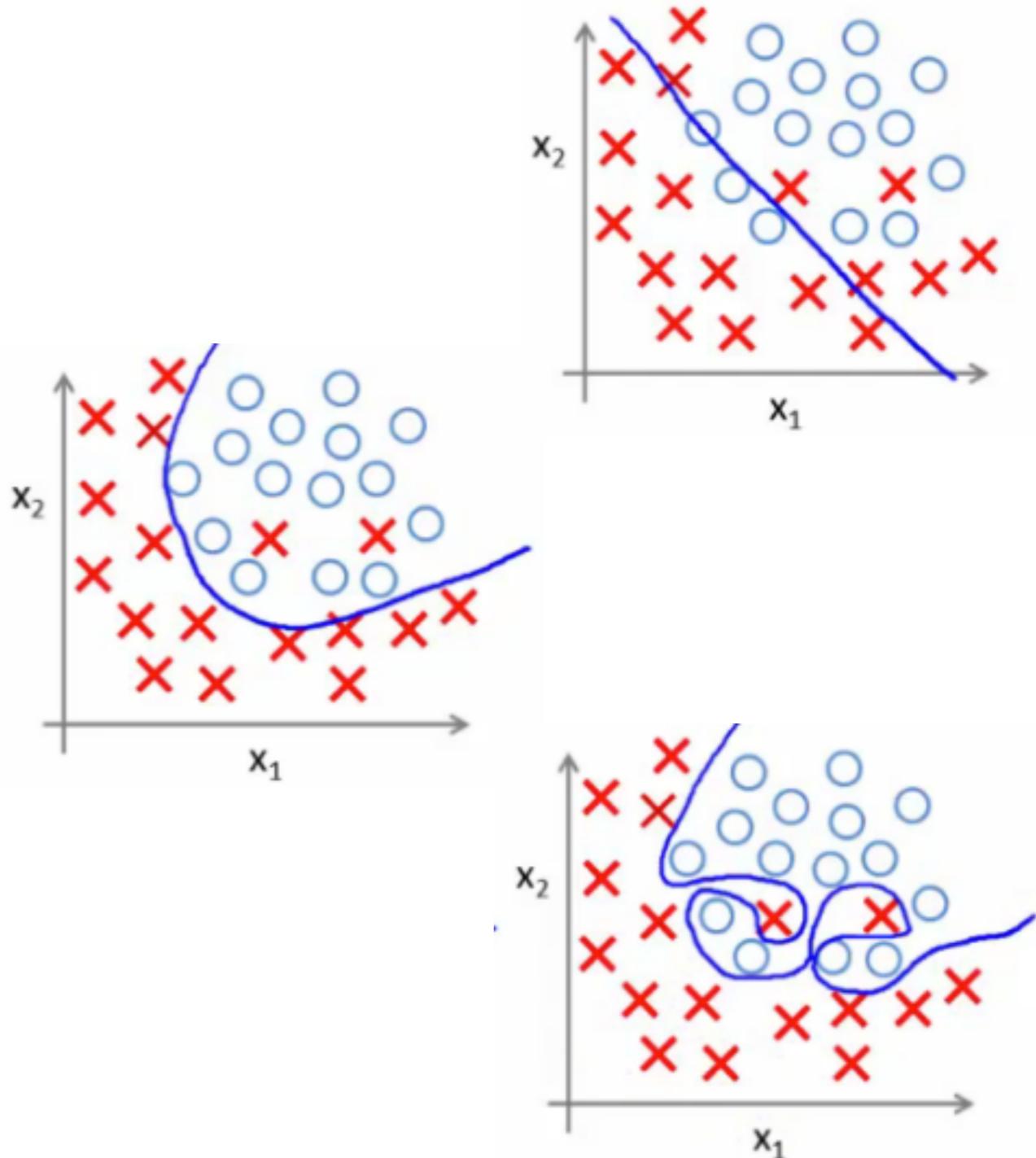
# Best model

- How resolve Underfitting?

- Check Errors

- How resolve Overfitting

- ?



# Train-Test Split

# Train-Test Split

- Modelling of new elements

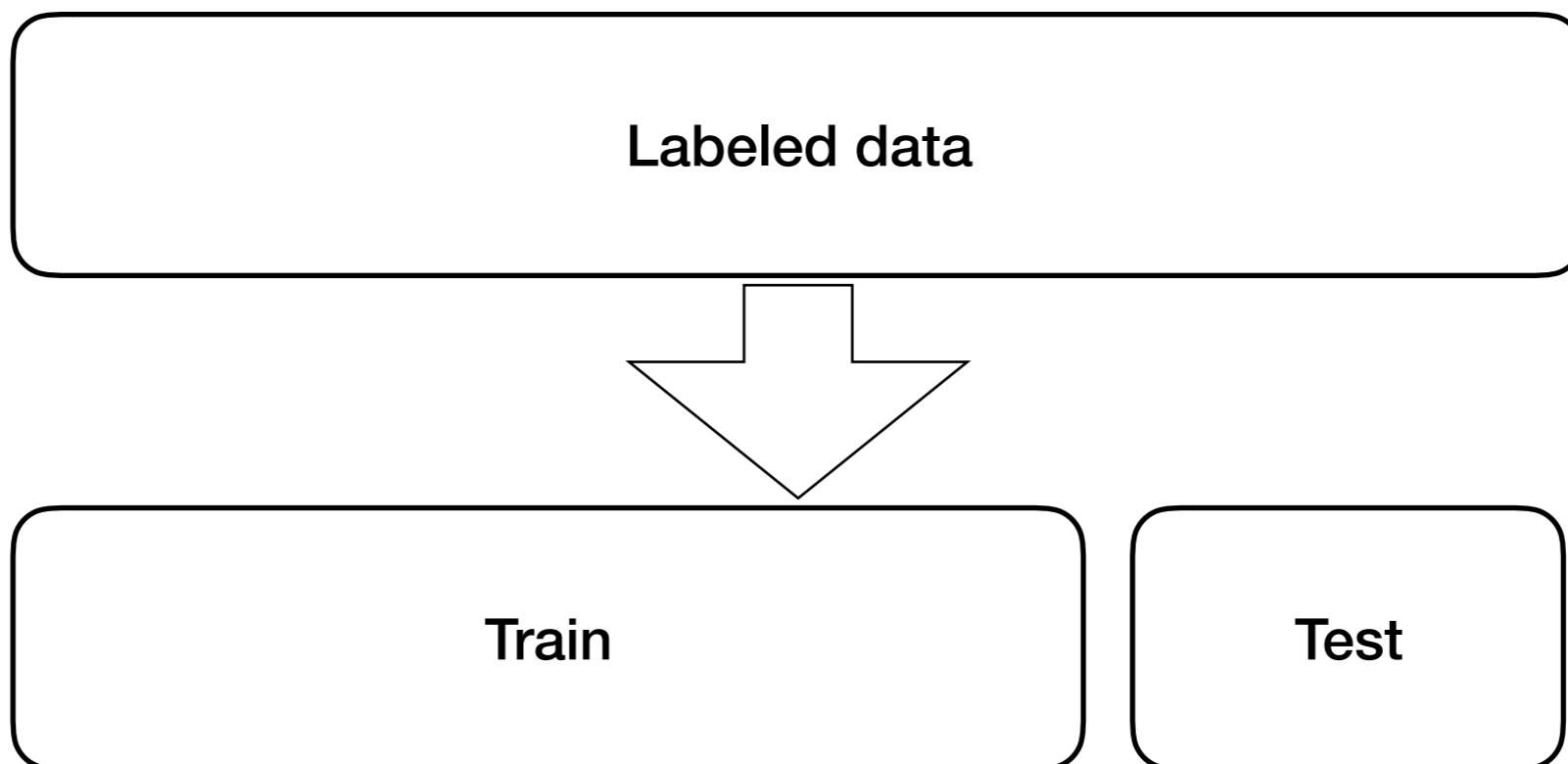
# Train-Test Split

- Modelling of new elements

Labeled data

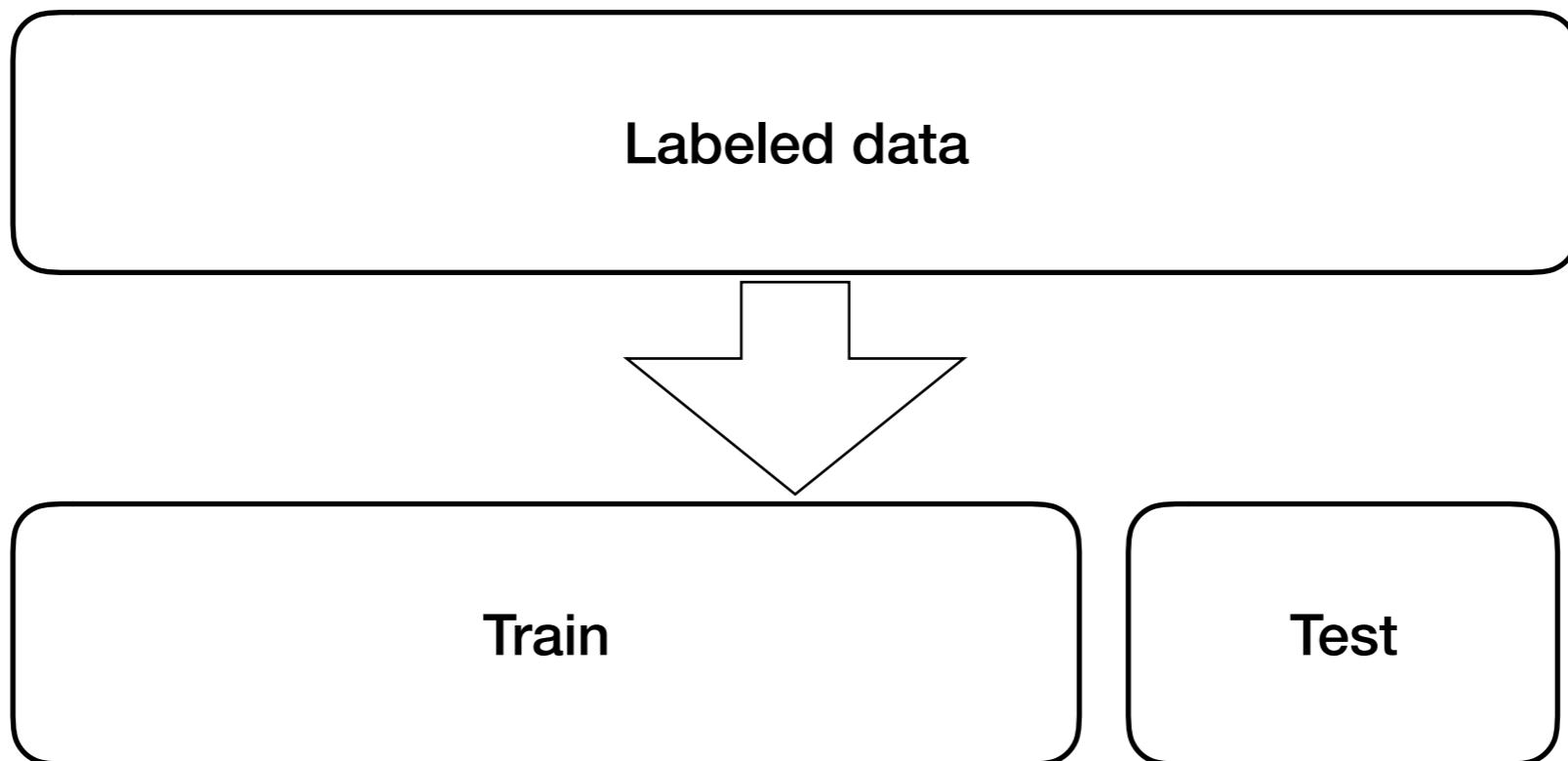
# Train-Test Split

- Modelling of new elements



# Holdout sample

- Modelling of new elements

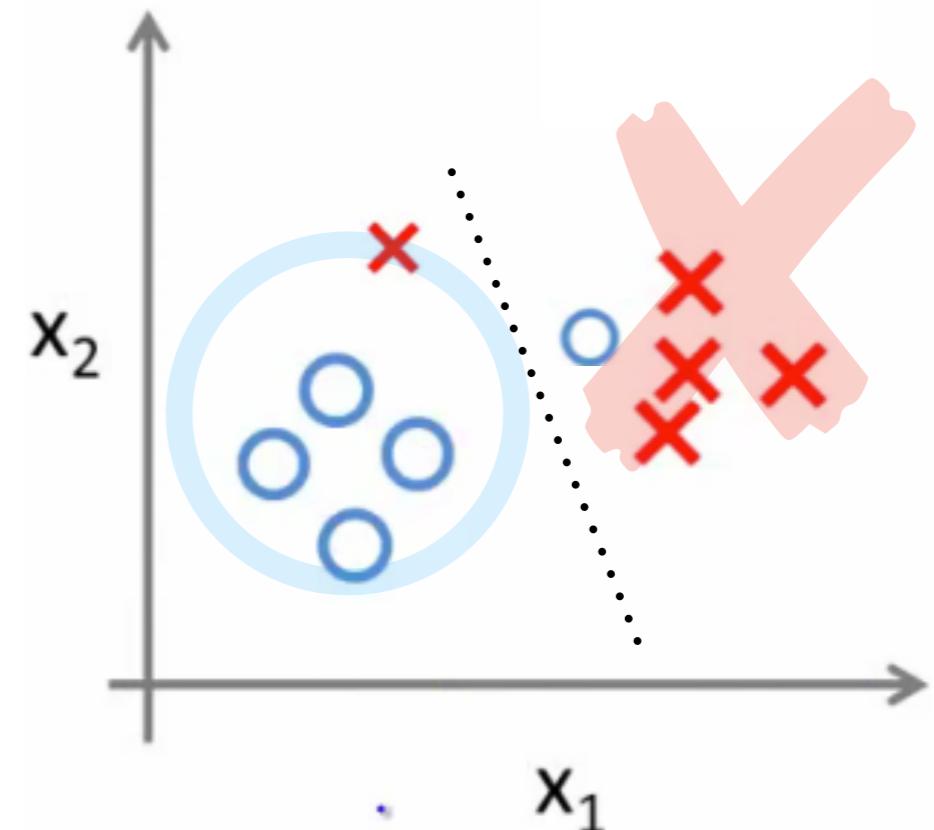


# K-fold Cross Validation

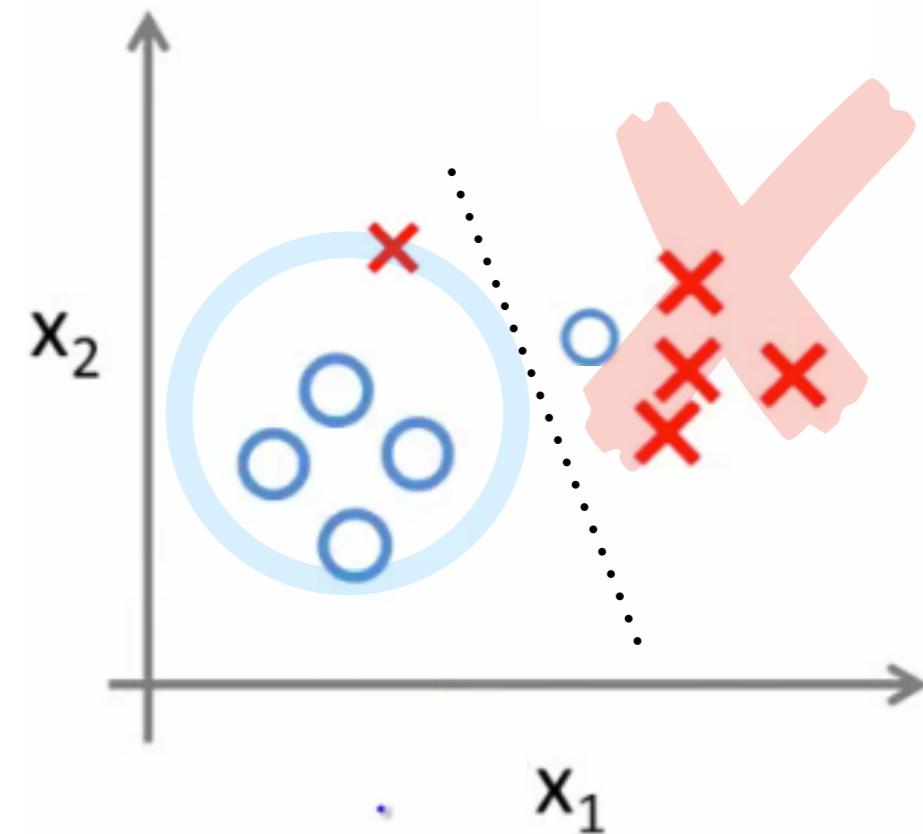
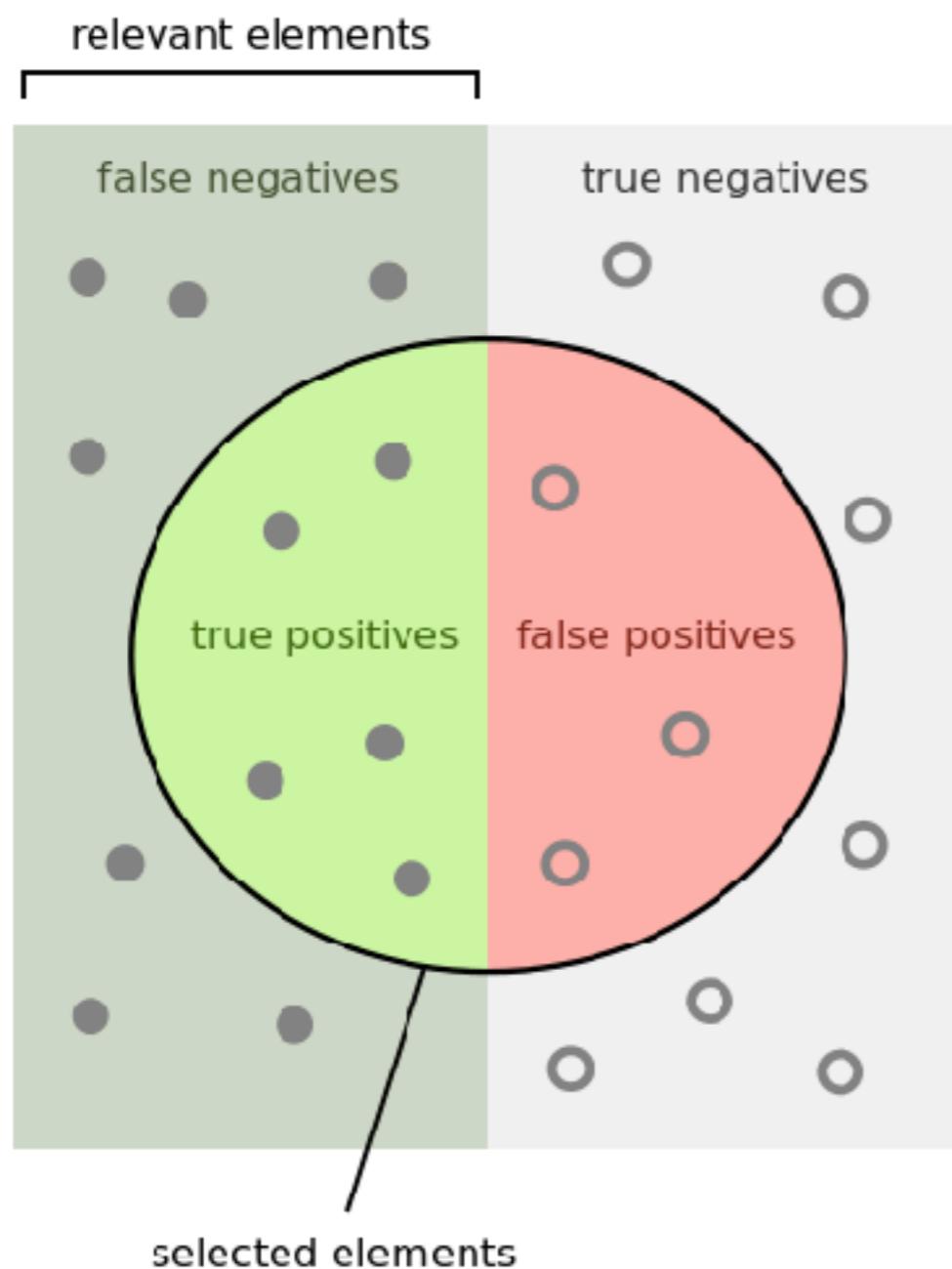


# Classification metrics

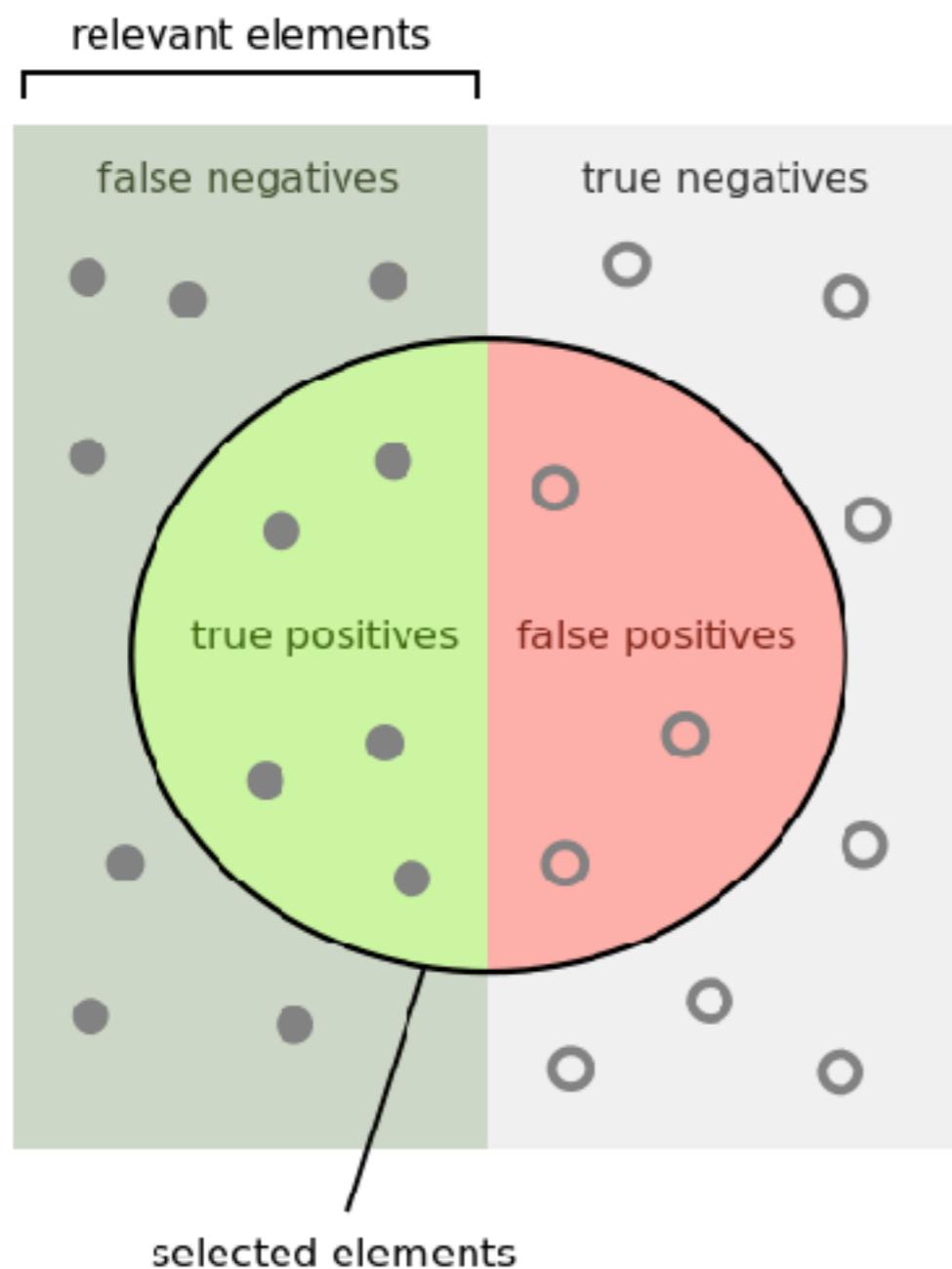
		Prediction outcome	
		P	n
actual value	p'	True Positive	False Negative
	n'	False Positive	True Negative



# Classification metrics



# Classification metrics



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$



How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$



$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

# Classification metrics

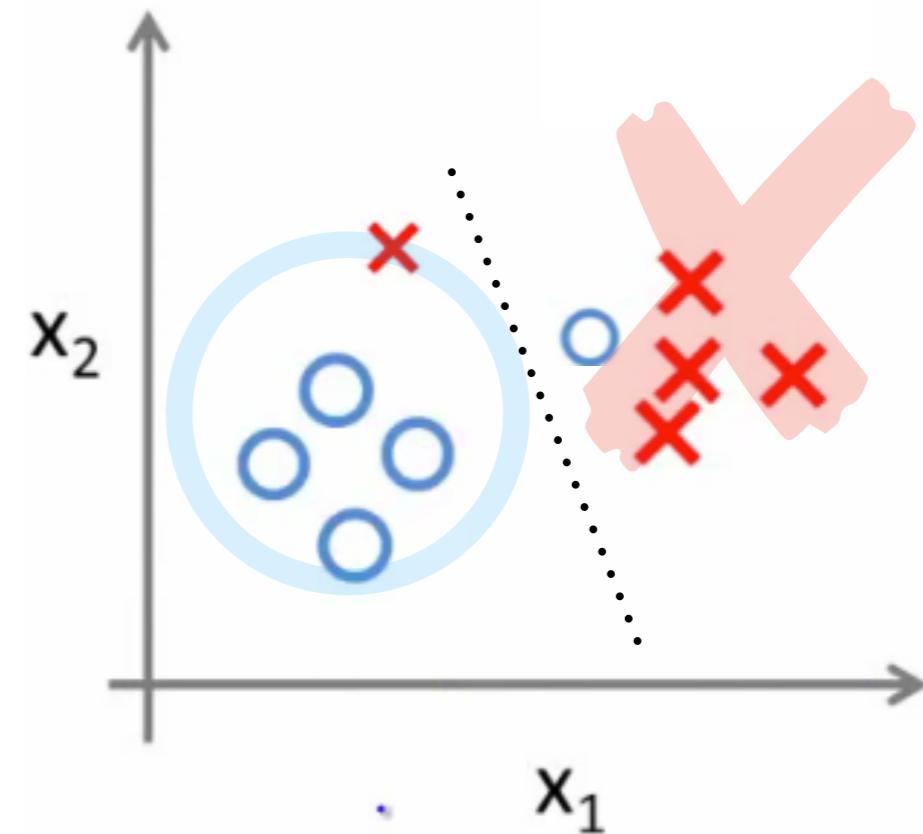
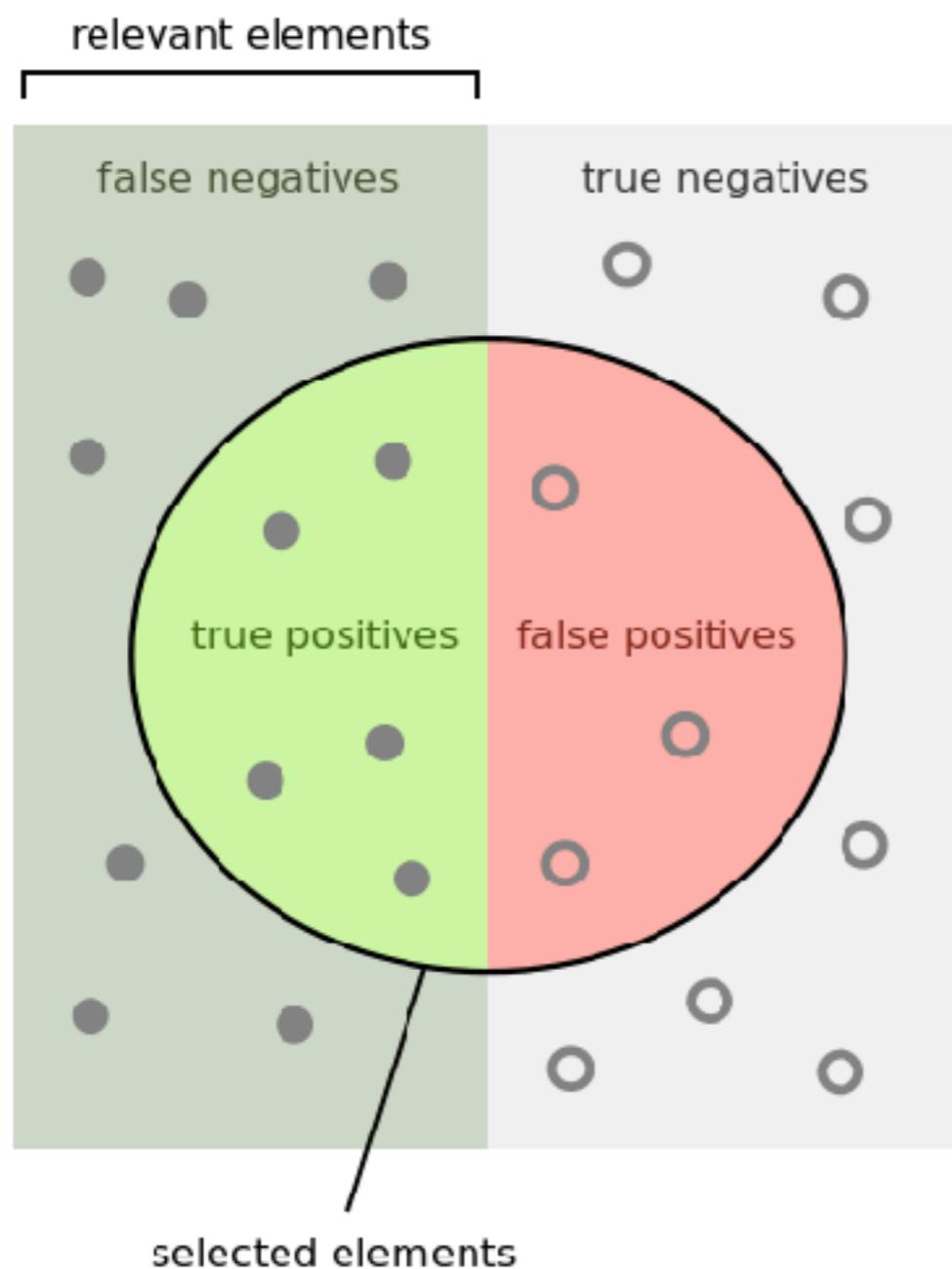
- Combine **precision** and **recall**
- F1-score  
(harmonic average of the precision and recall)

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

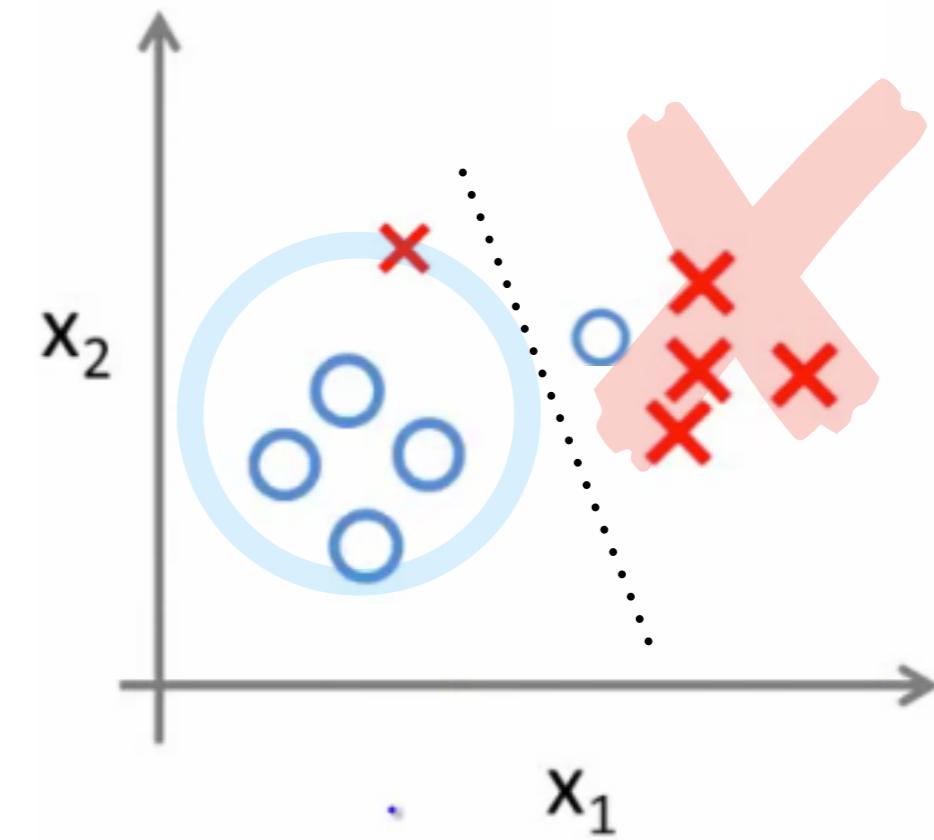
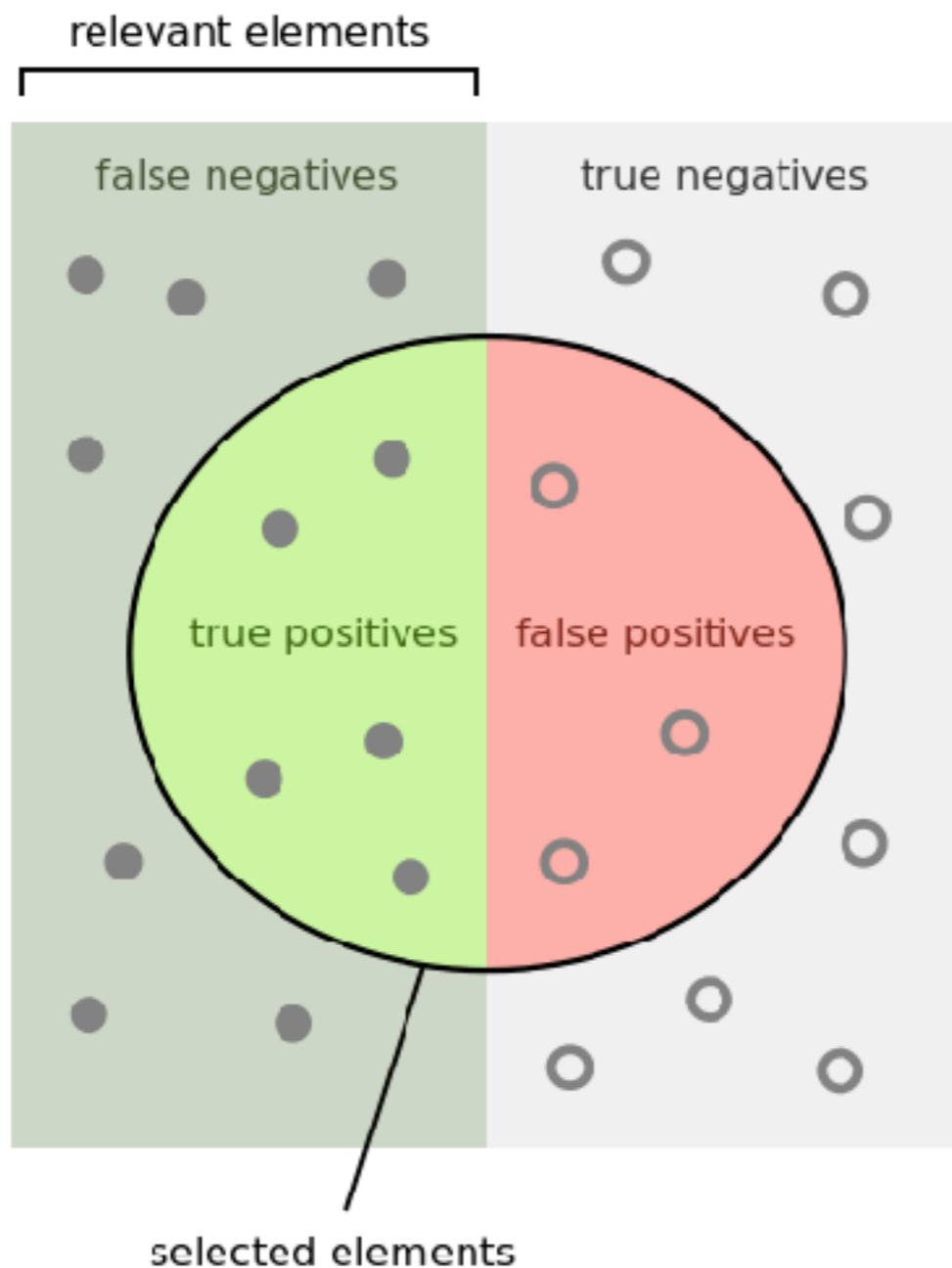
- F $\beta$ -score

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}$$

# Classification metrics

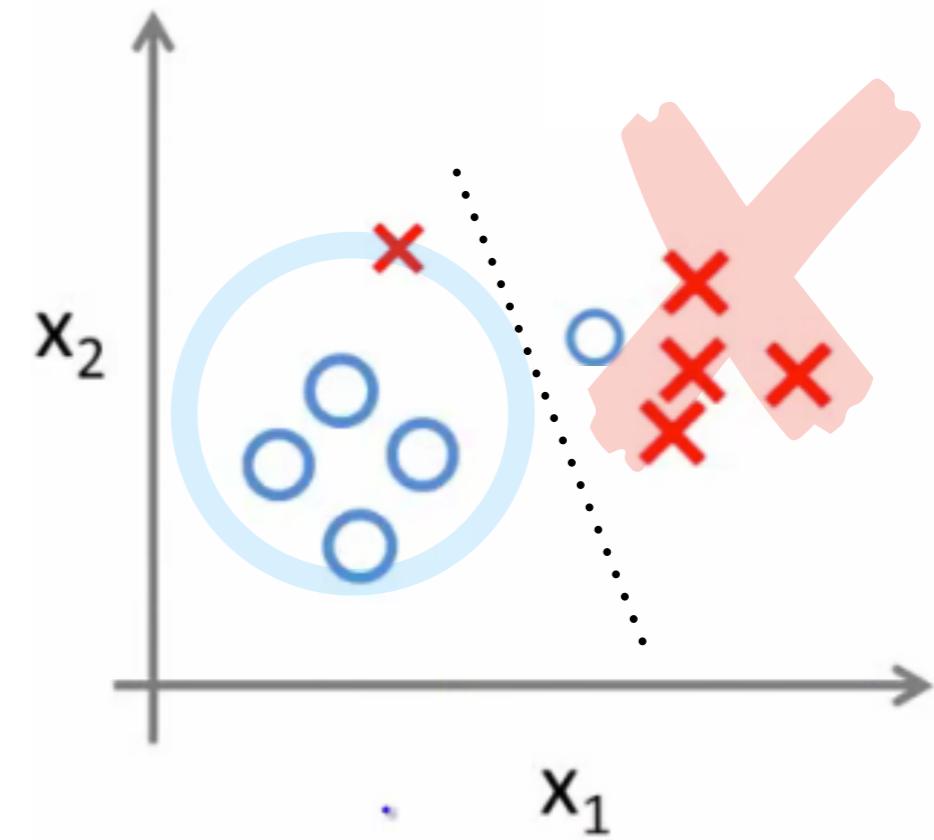
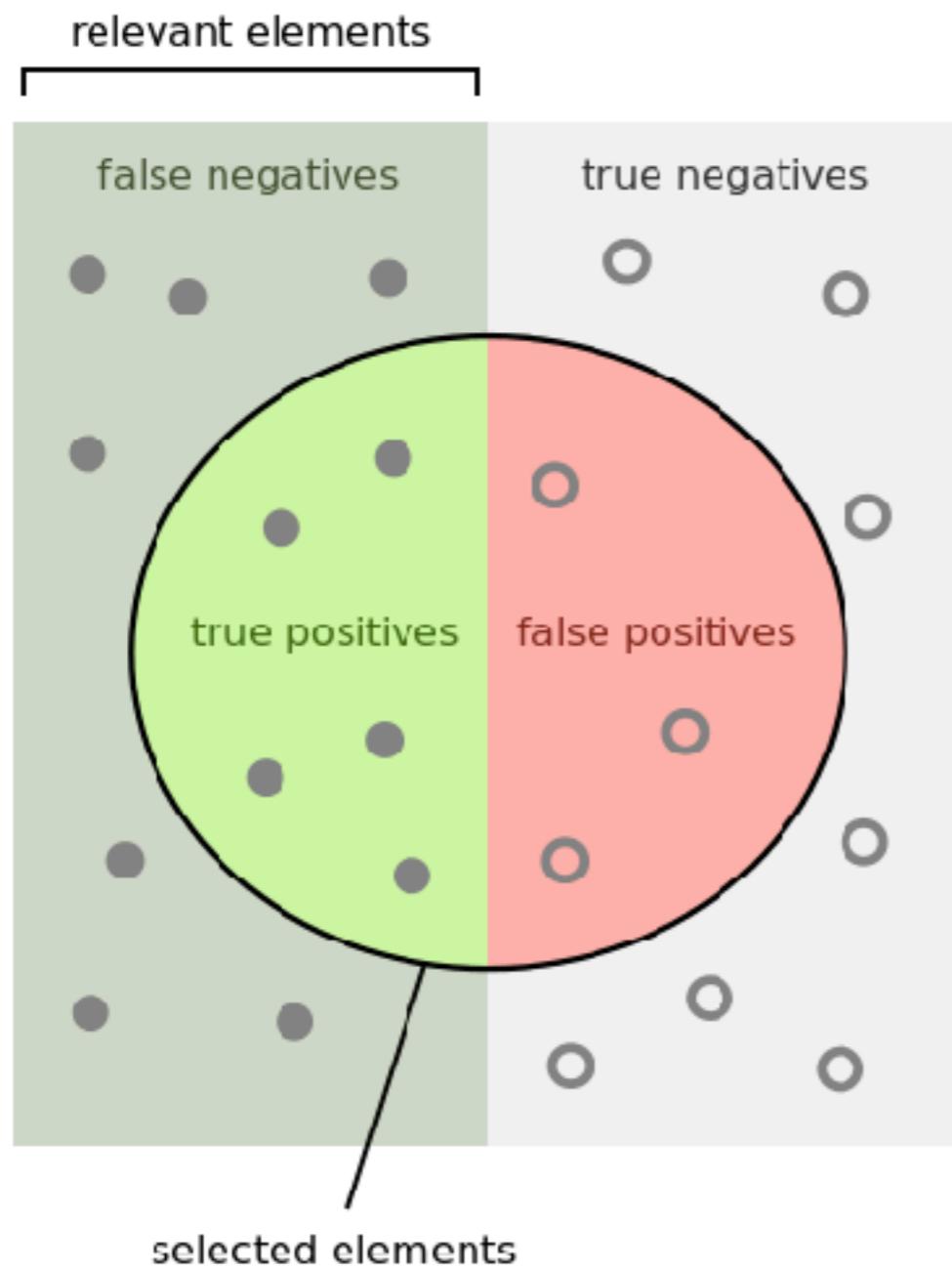


# Classification metrics



$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

# Classification metrics



$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{TN + FP}$$

# Confusion matrix

		Prediction outcome	
		P	n
		True Positive	False Negative
actual value	p'	True Positive	False Negative
	n'	False Positive	True Negative

Actual class	Predicted class	
	Cat	Non-cat
Cat	5 True Positives	2 False Negatives
Non-cat	3 False Positives	17 True Negatives

# Confusion matrix

		Actual class		
		Cat	Dog	Rabbit
Predicted class	Cat	5	2	0
	Dog	3	3	2
	Rabbit	0	1	11

# Classification metrics

- True positive, False negative, True negative, False positive
- Precision, Recall
- F1-score/F $\beta$ -score
- Accuracy
- True positive rate, False positive rate