Introduction to ML

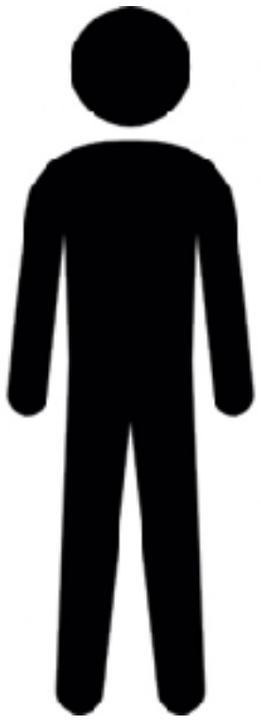
Ermakov Petr
17.10.2017

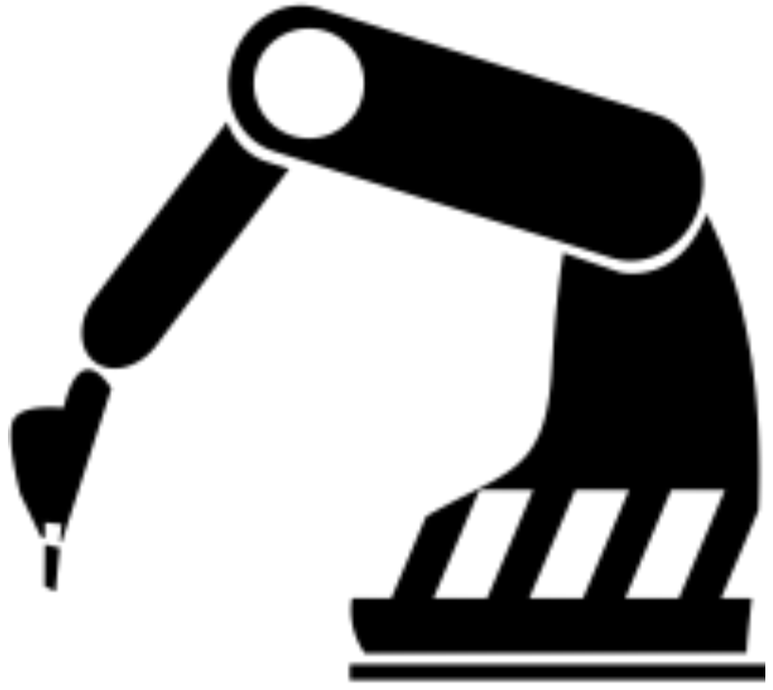
About me

- Ermakov Petr
- 2016 - until now: Head of the Analytics Department (Mail.Ru, Youla)
- 2013 - 2016: Senior Data Scientist (HH.ru)

ermakov.pd@gmail.com
facebook.com/ermakovpetr
linkedin.com/in/ermakovpetr

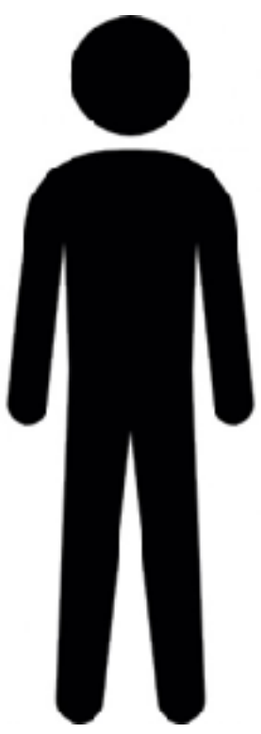
Machine Learning and Human



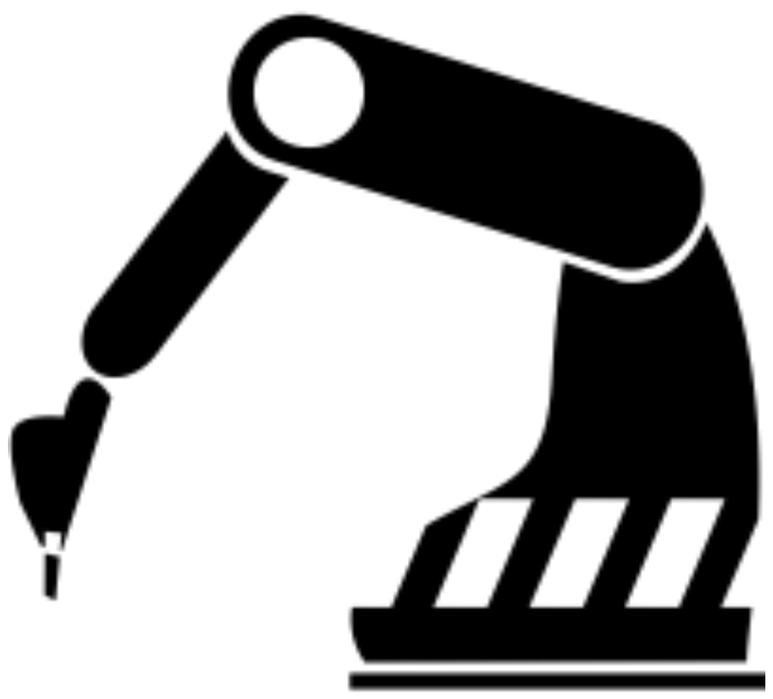


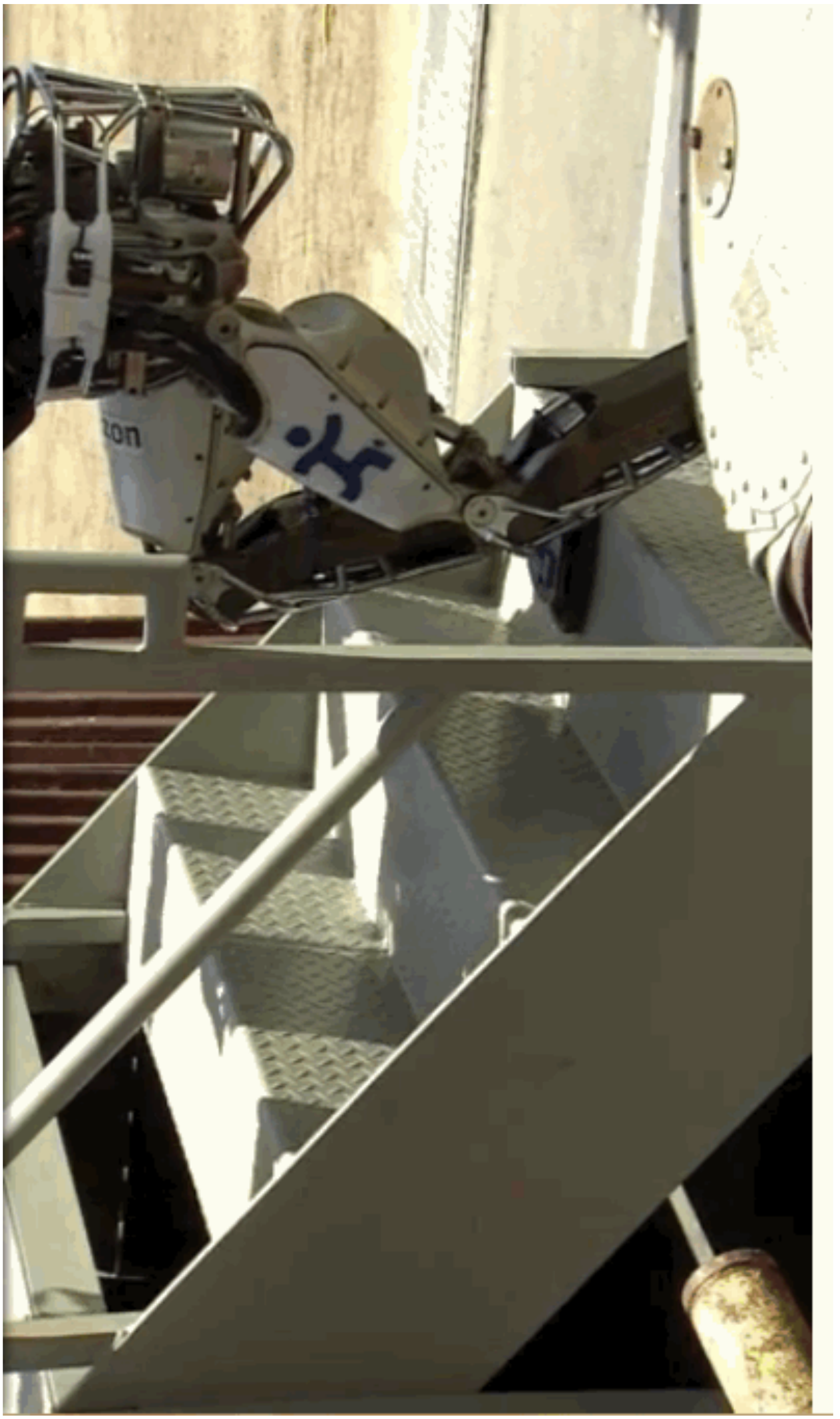
Machine Learning and Human

Learn from experience



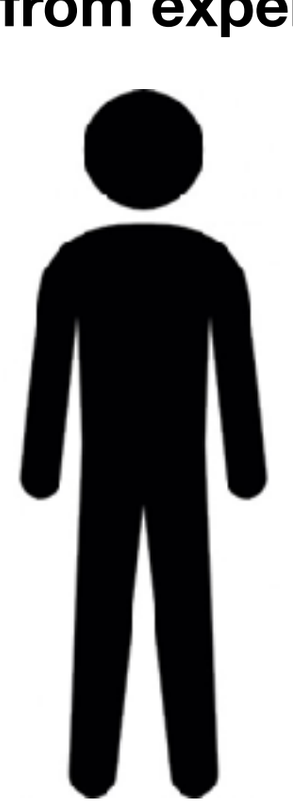
Follow instruction

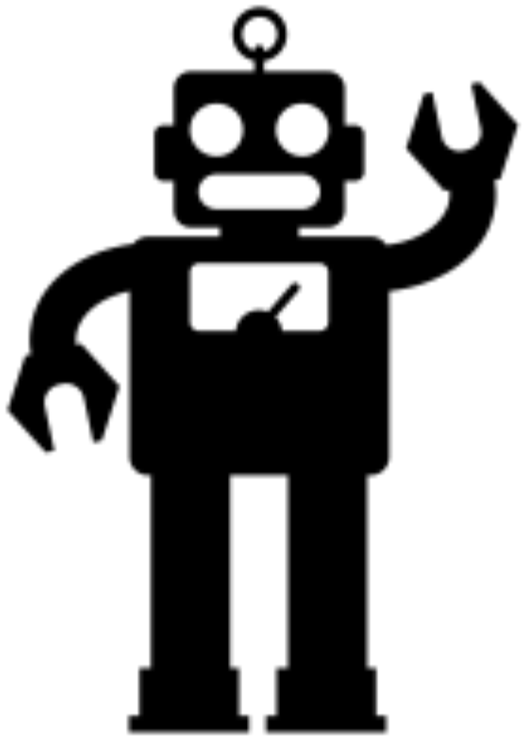




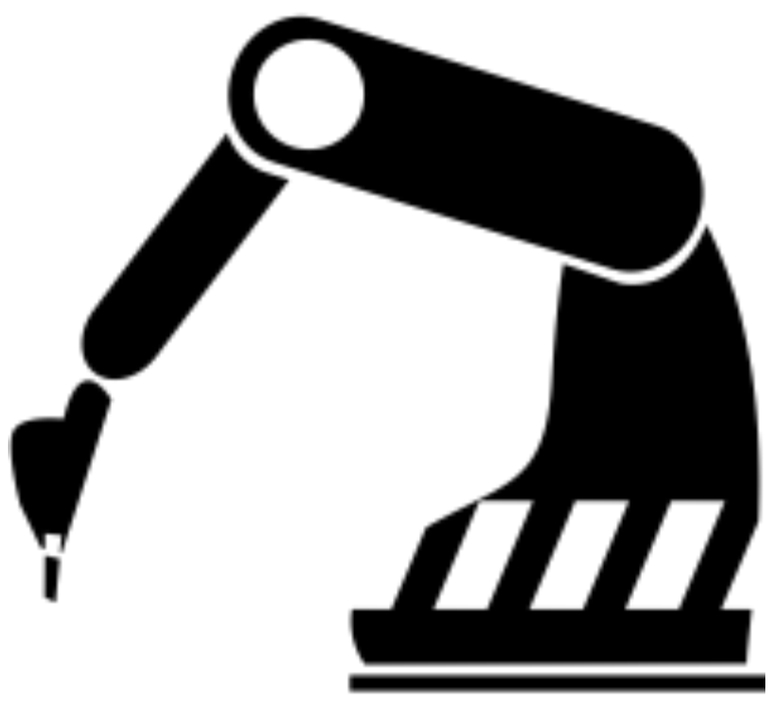
Machine Learning and Human

Learn from experience



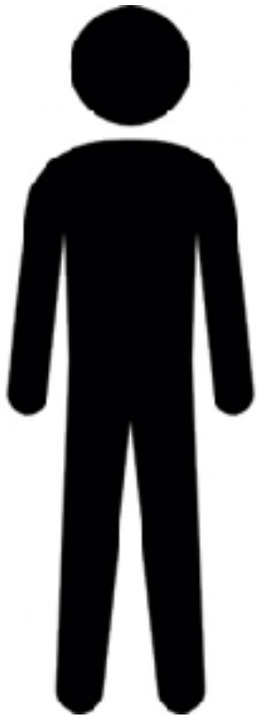


Follow instruction

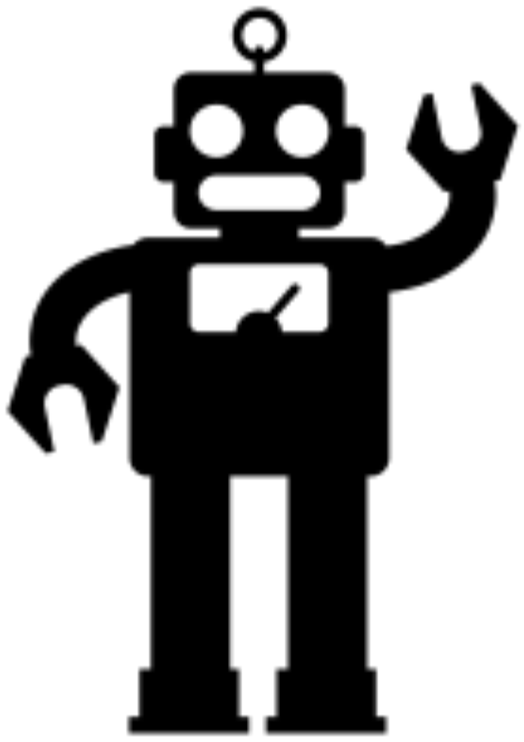


Machine Learning and Human

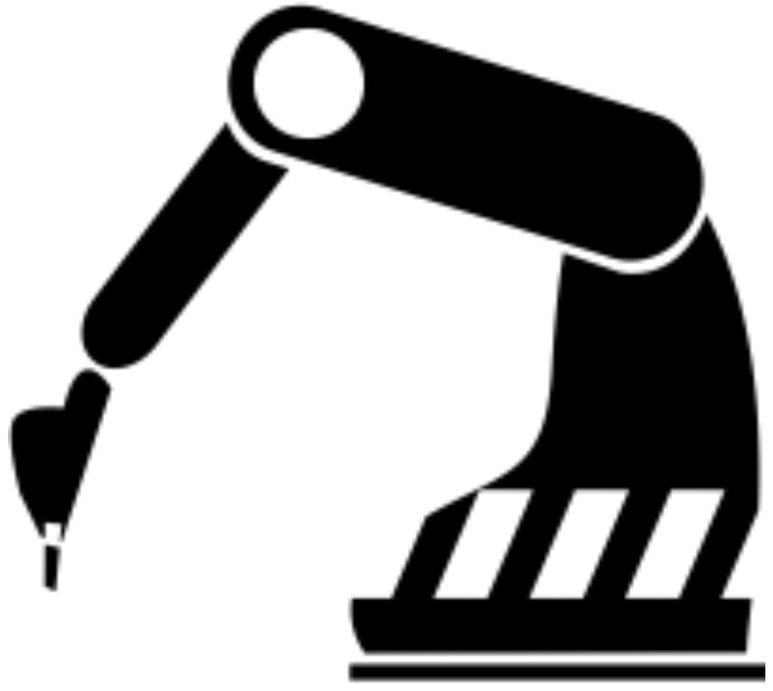
Learn from experience



Learn from experience



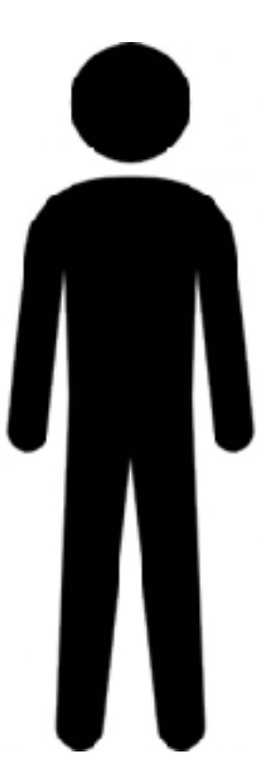
Follow instruction

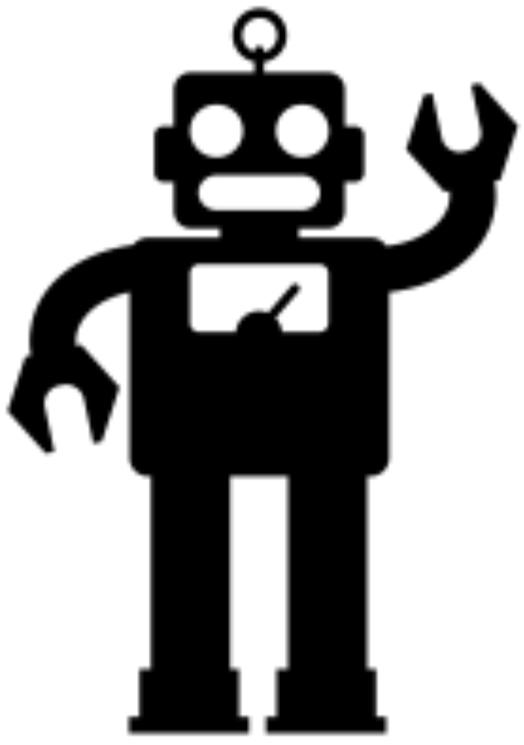


Machine Learning and Human

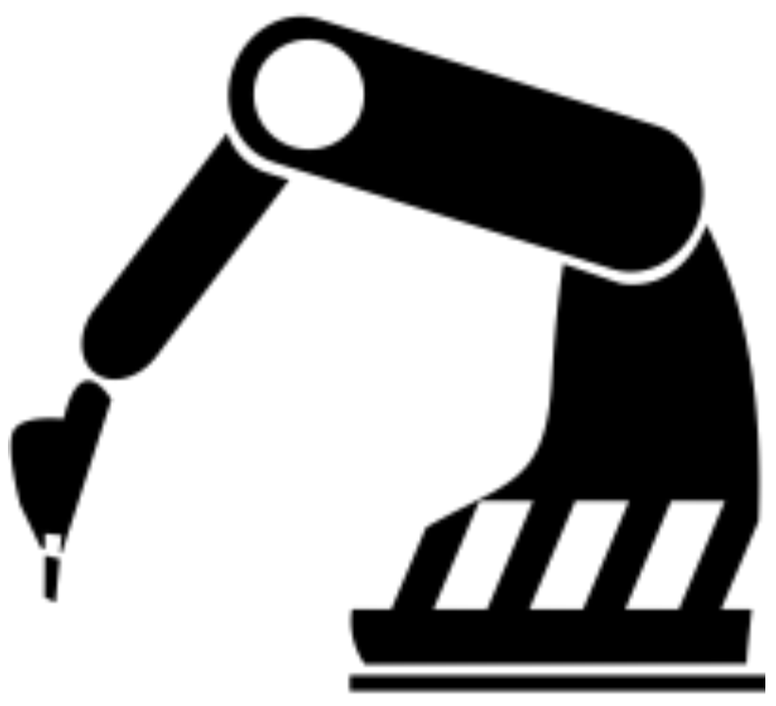
Learn from ~~experience~~ data

Learn from experience





Follow instruction



Machine Learning

Artificial Intelligence

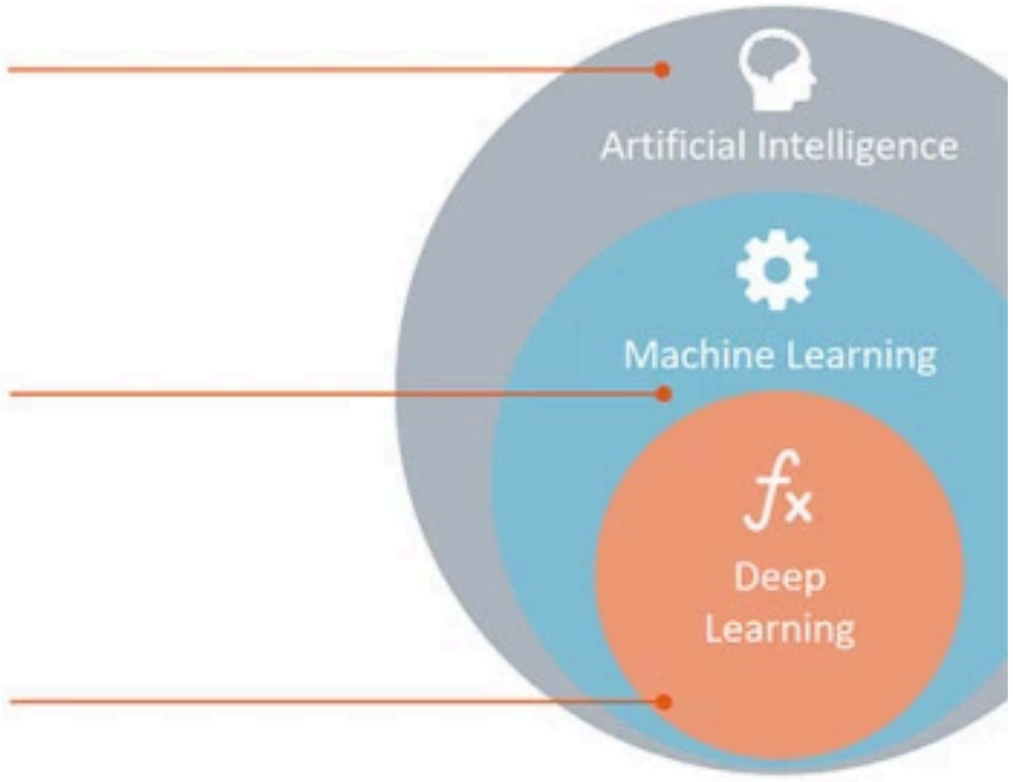
Any technique which enables computers to mimic human behavior.

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

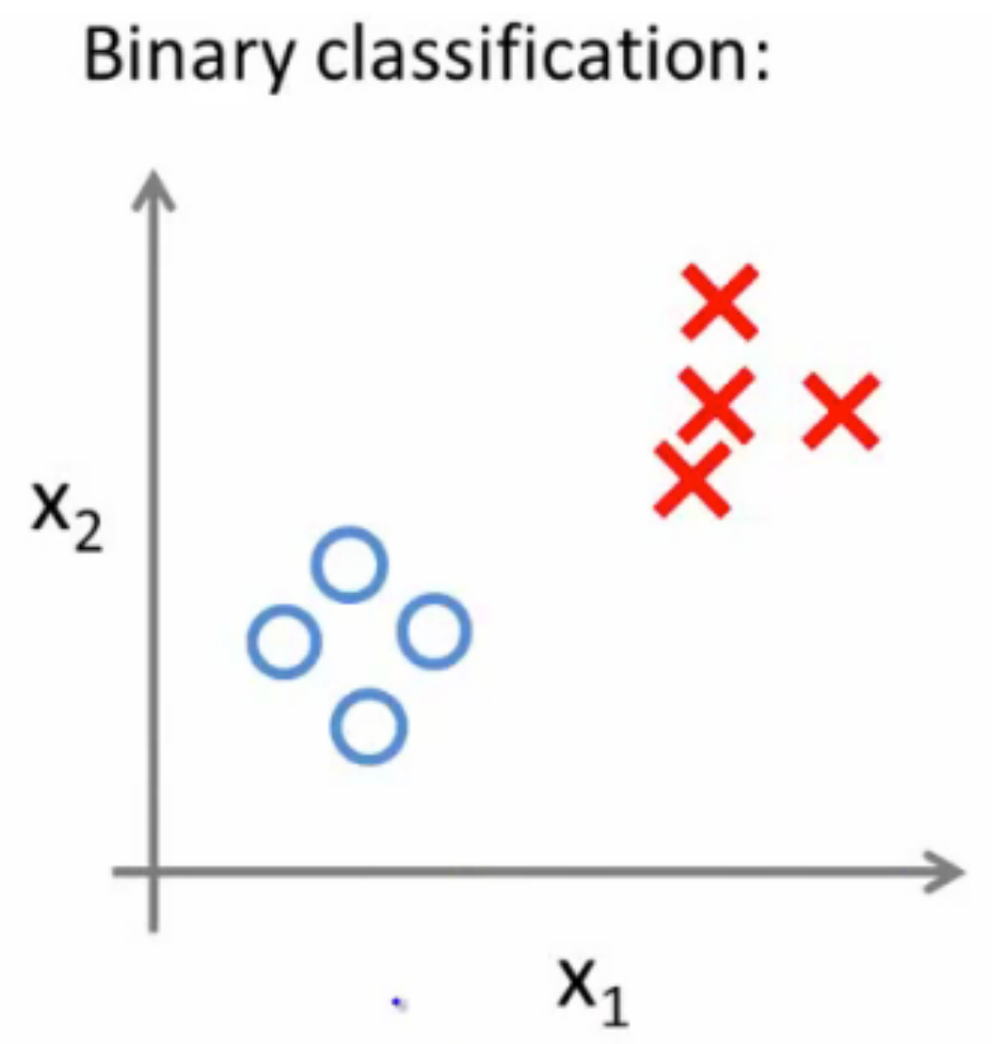

MARIO
525150
×42
WORLD
8-2
TIME
364

Supervised learning

- Classification
- Regression
- … other …

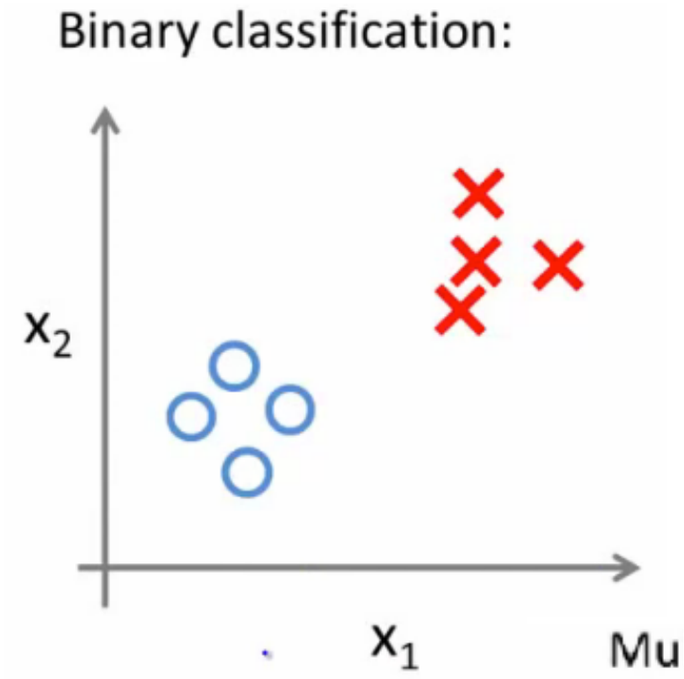
Classification

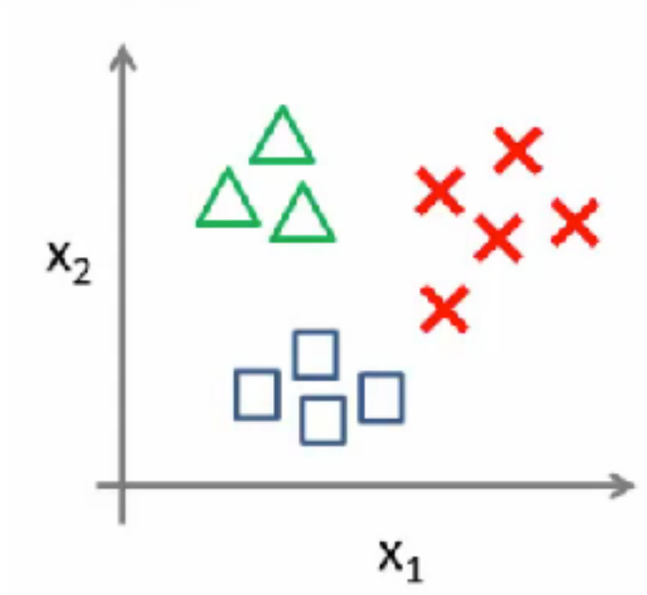
- Binary classification
    - Two classes



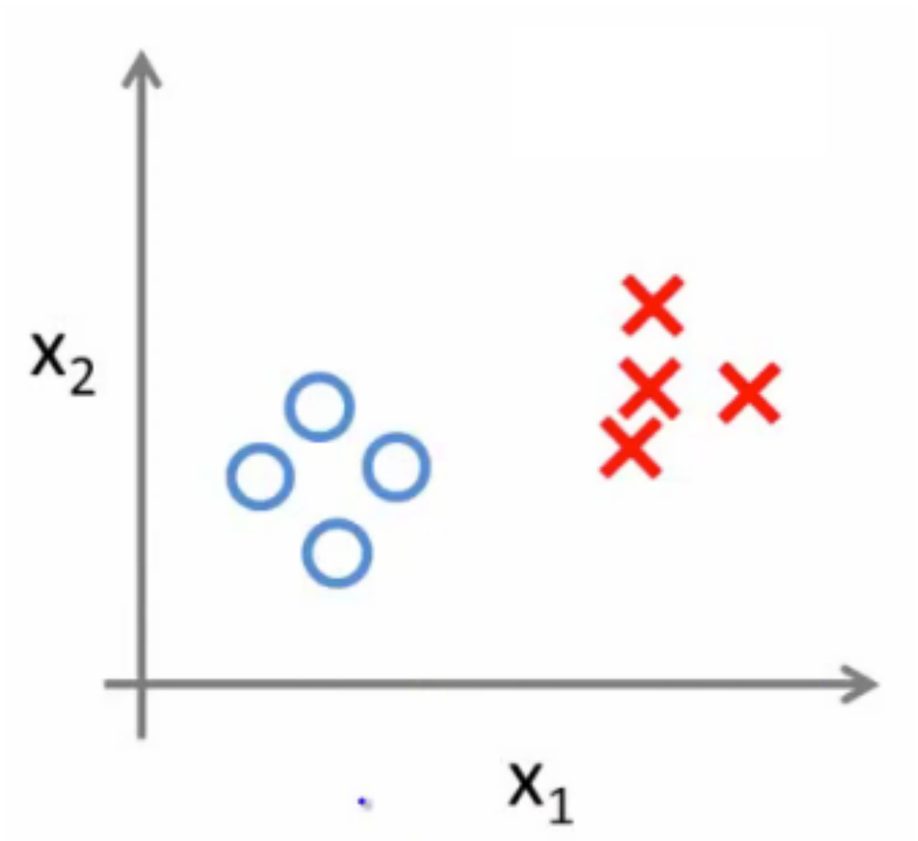
Classification

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

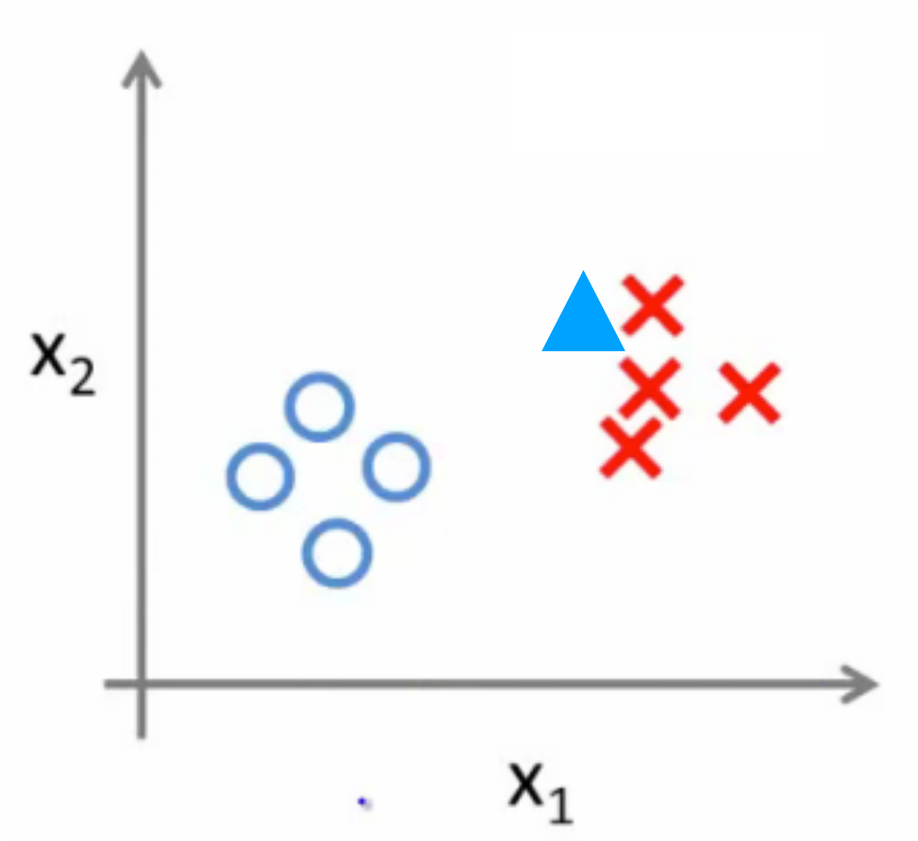




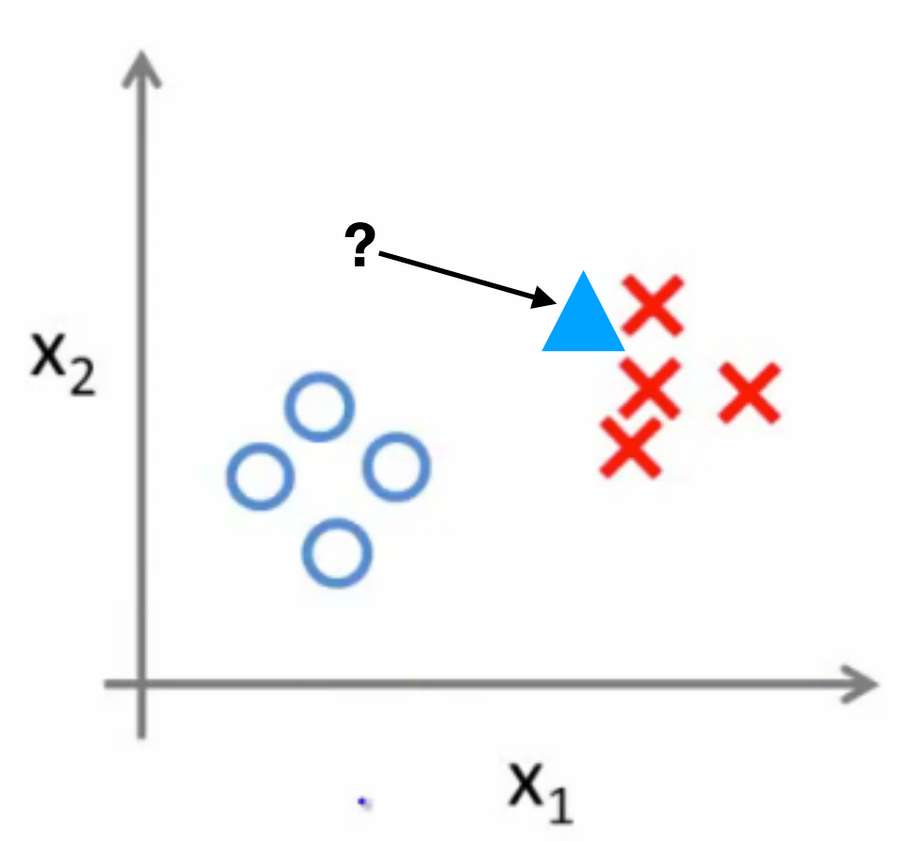
New element for classification



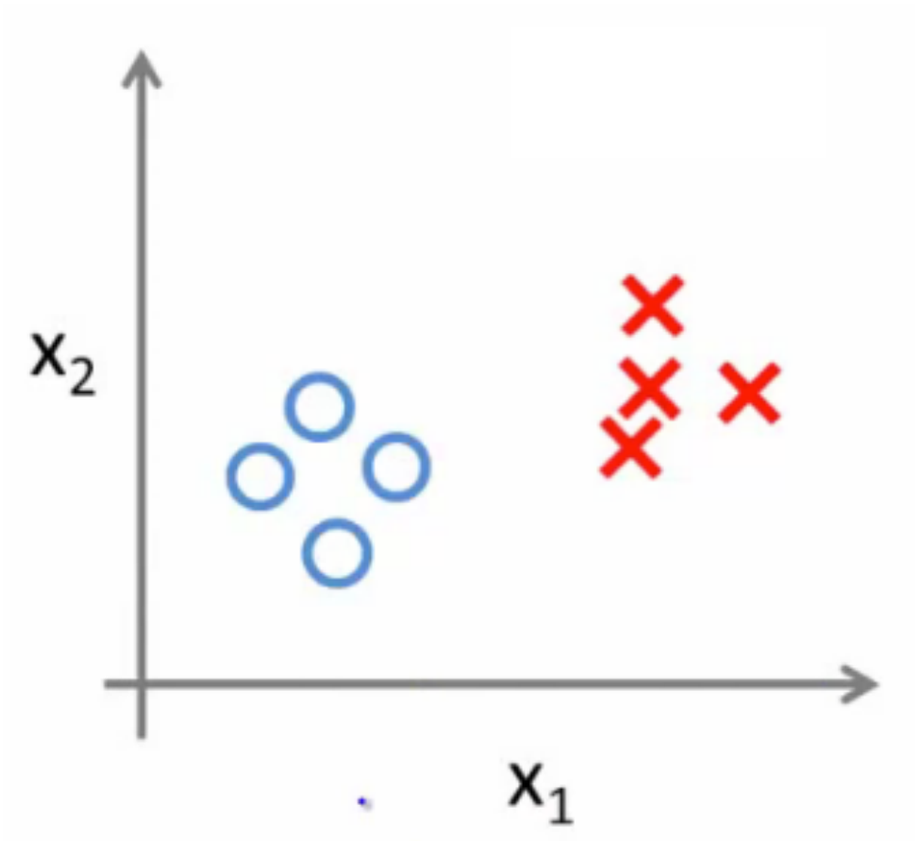
New element for classification



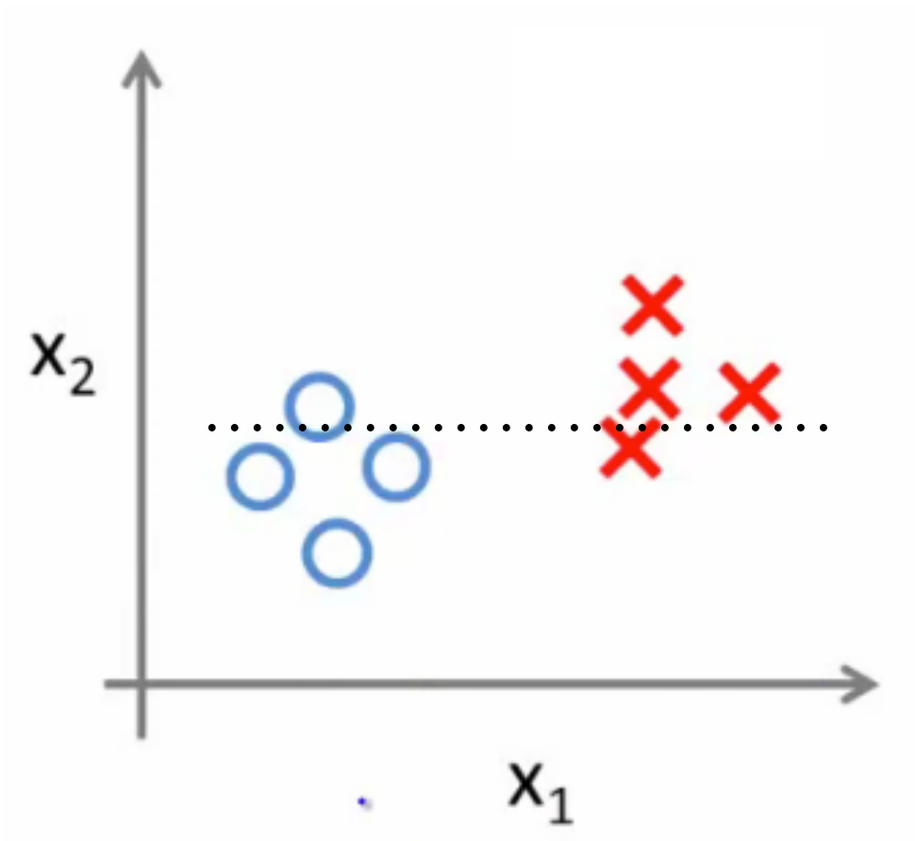
New element for classification



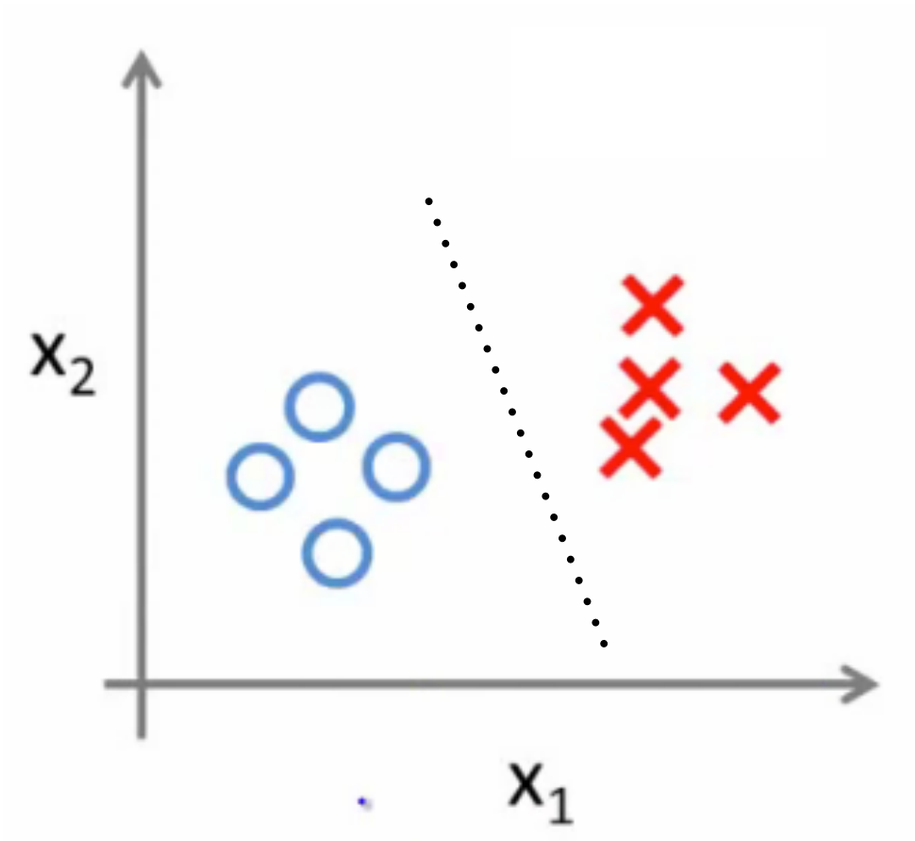
Build first model



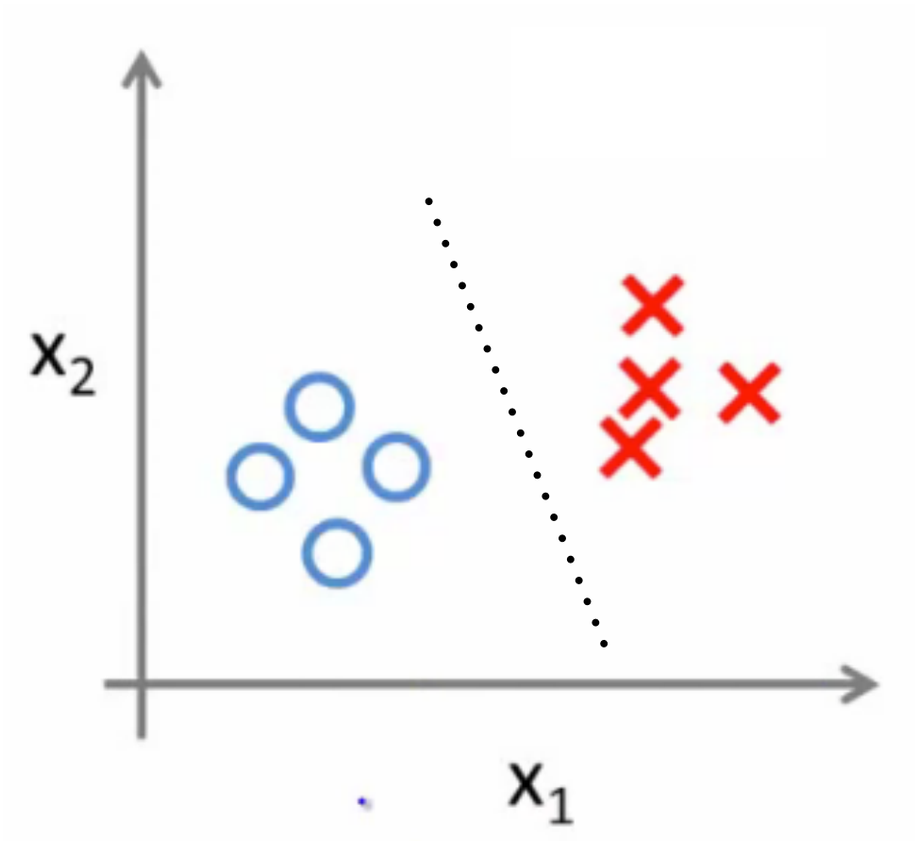
Build first model



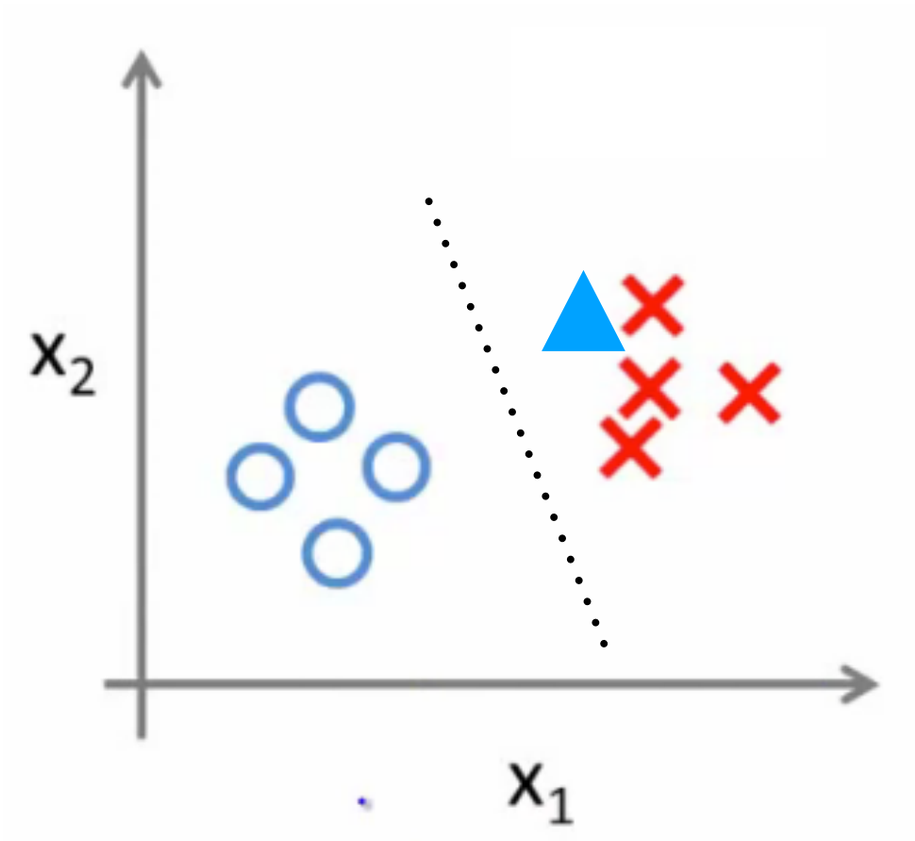
Build first model



Build first model



Build first model



Classification

- X - set of object descriptions (X1, X2) - features
- Y - set of classes (circles and crosses) - target
- a: X -> Y - ML algorithm

Titanic

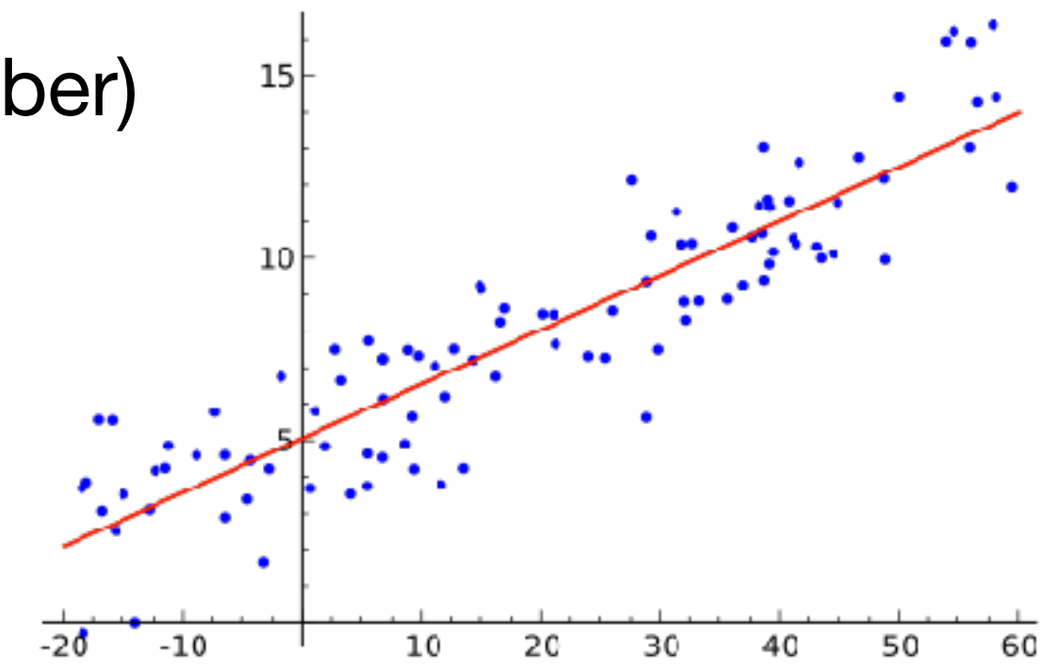
- Which passengers survived the tragedy?
- X (features):
  - Ticket class
  - Sex
  - Port of Embarkation
  - Ticket number
  - Age in years, …
- Y (target):
  - Survival

    0 = No, 1 = Yes

https://www.kaggle.com/c/titanic/

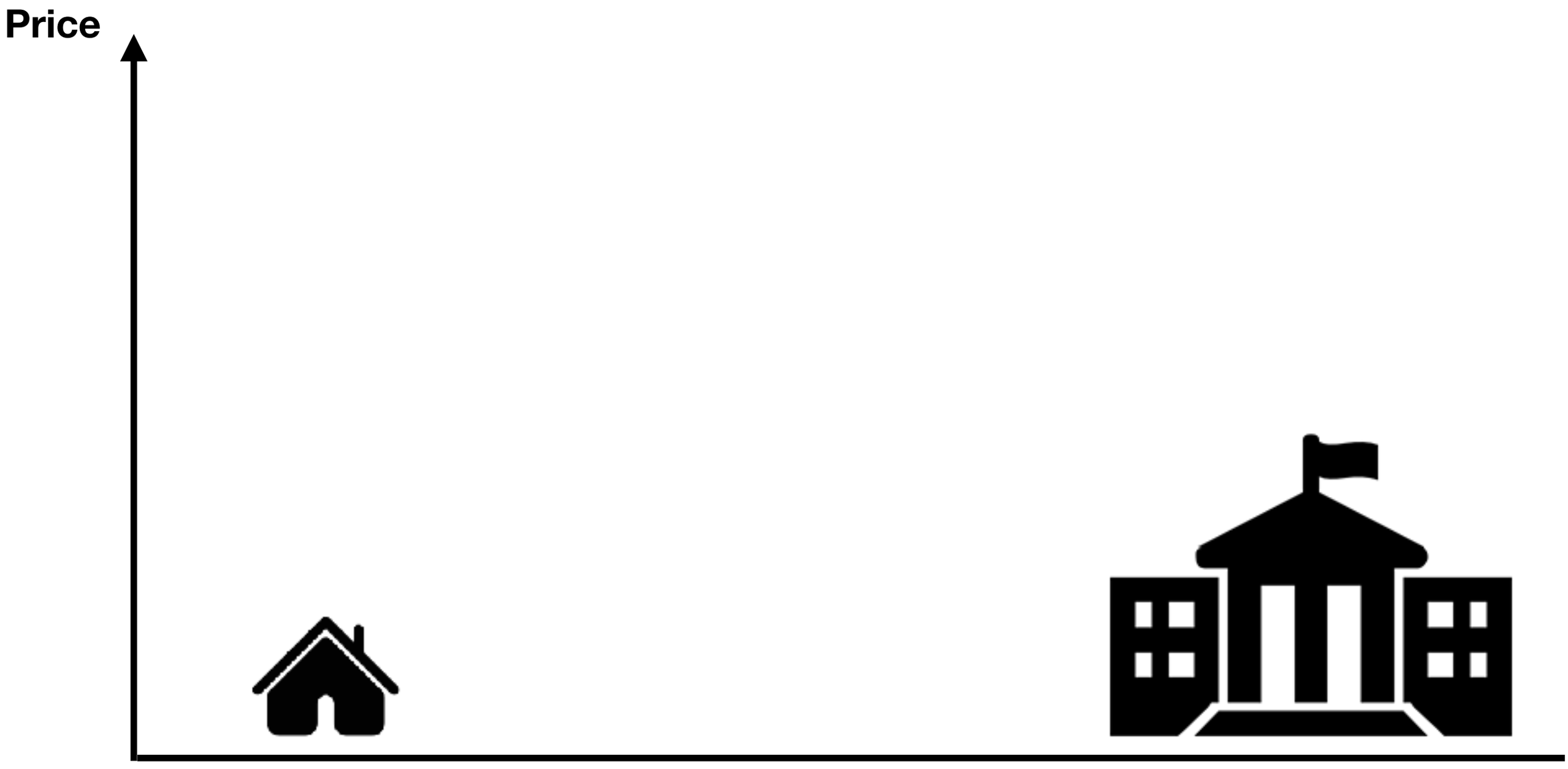
Regression

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

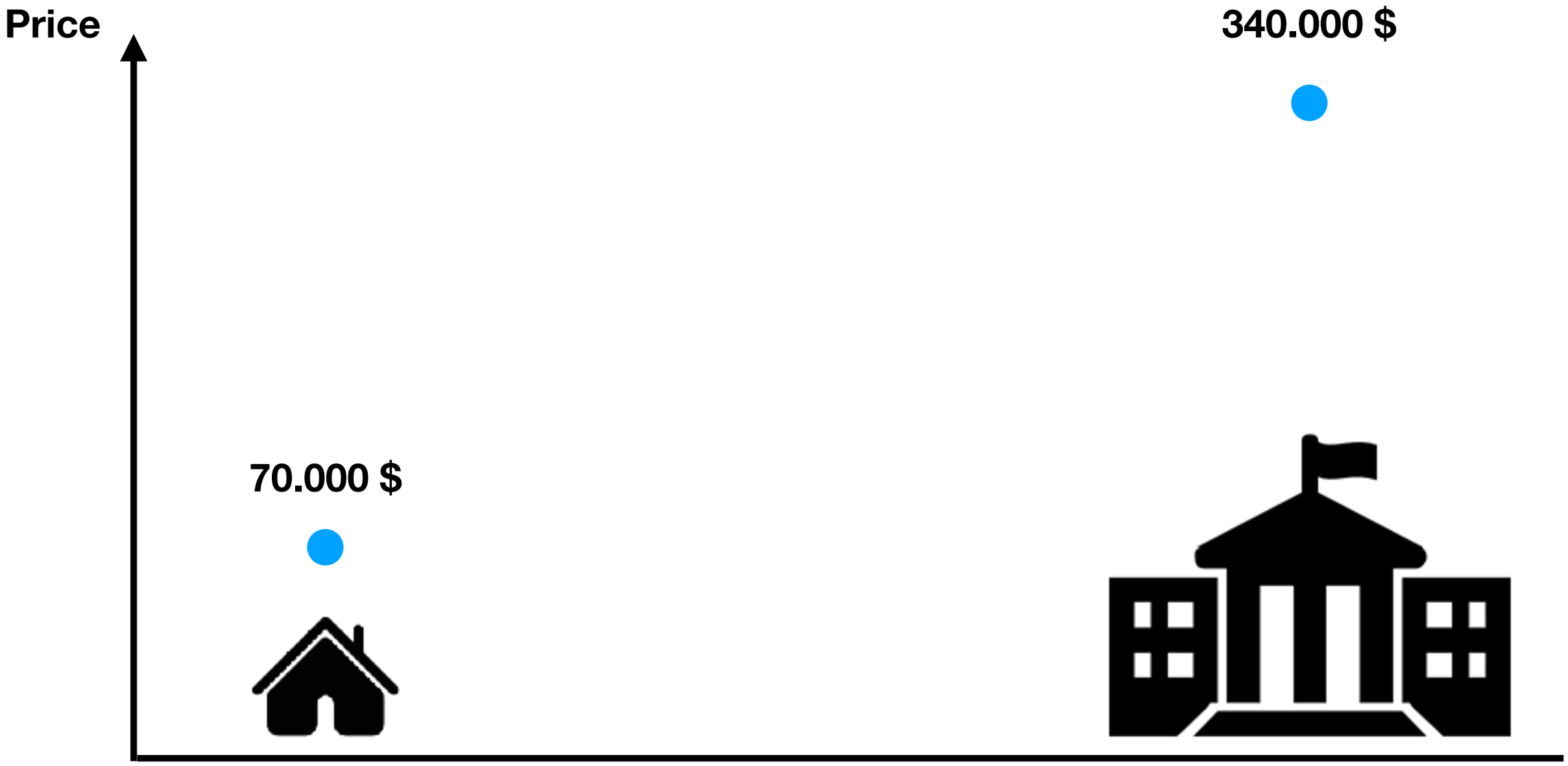


Regression

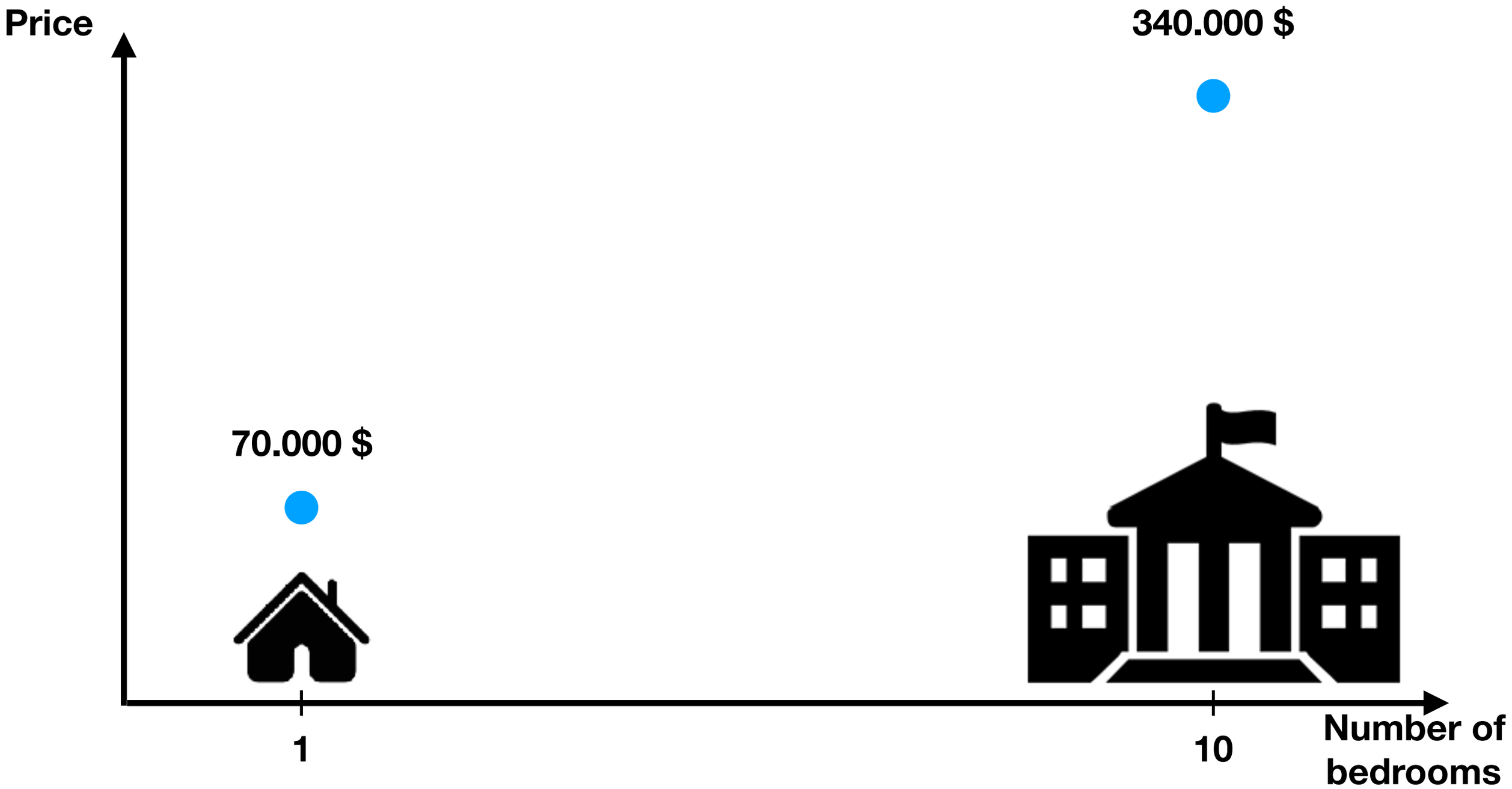
Regression



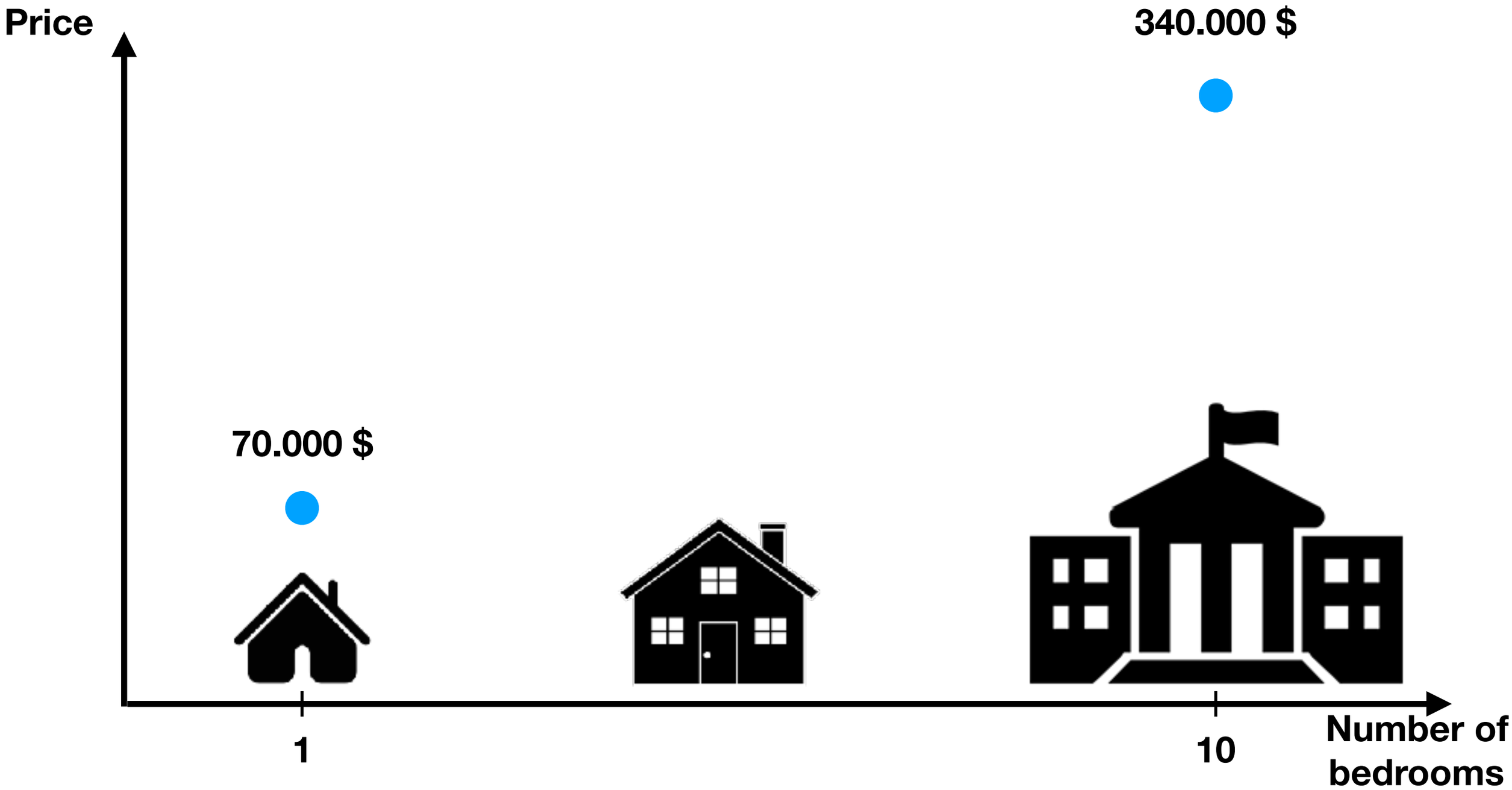
Regression



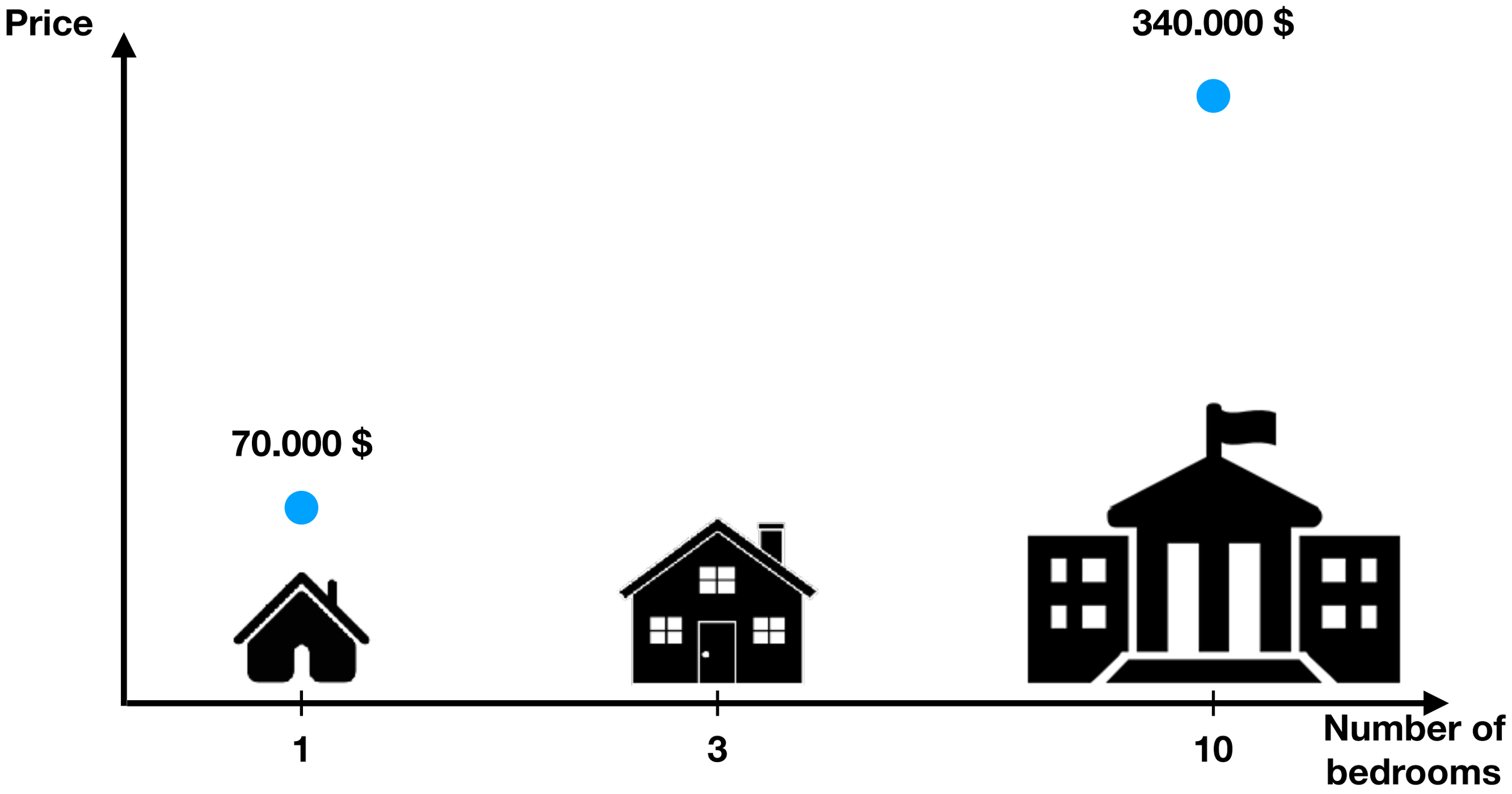
Regression



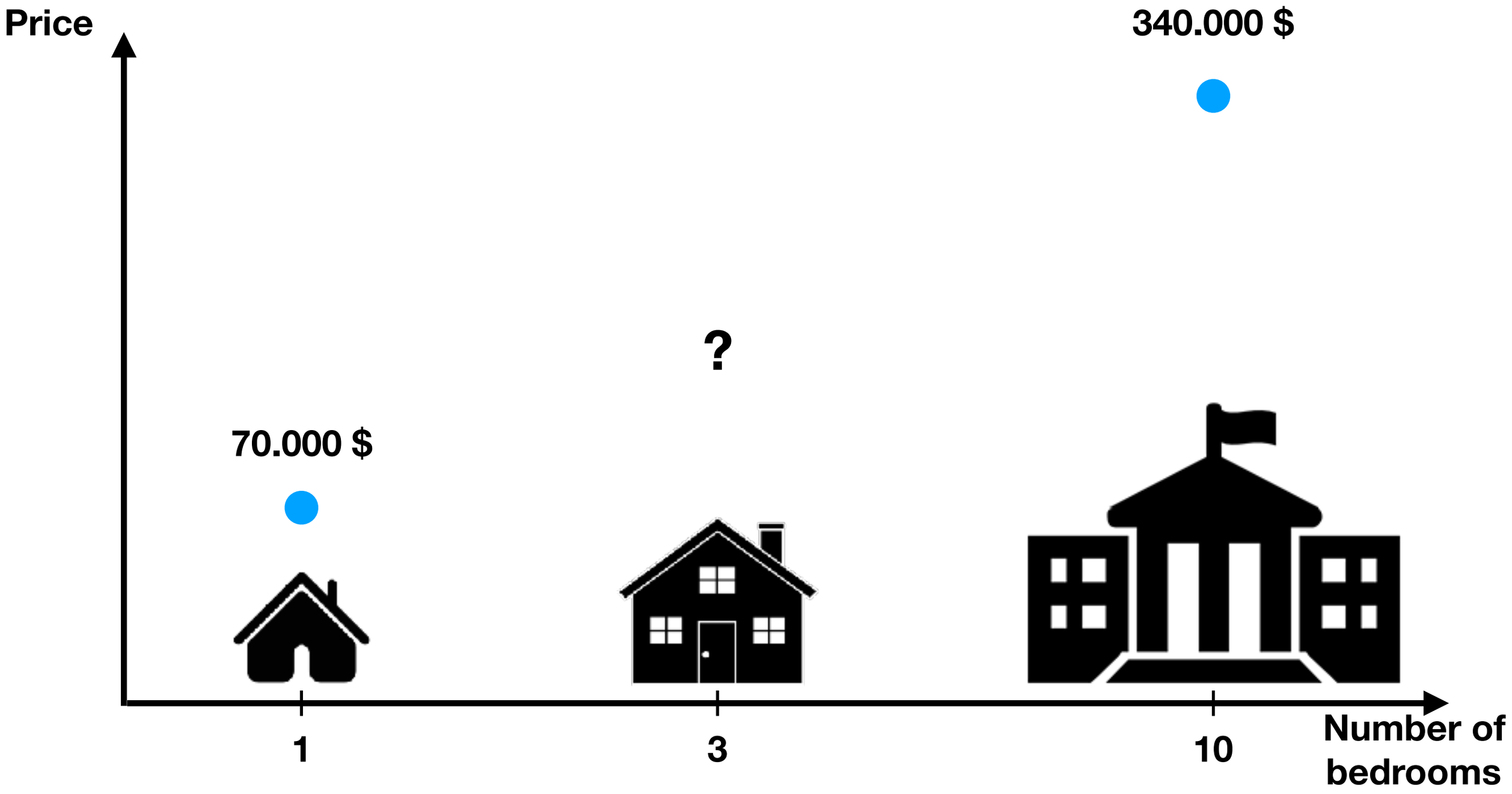
Regression



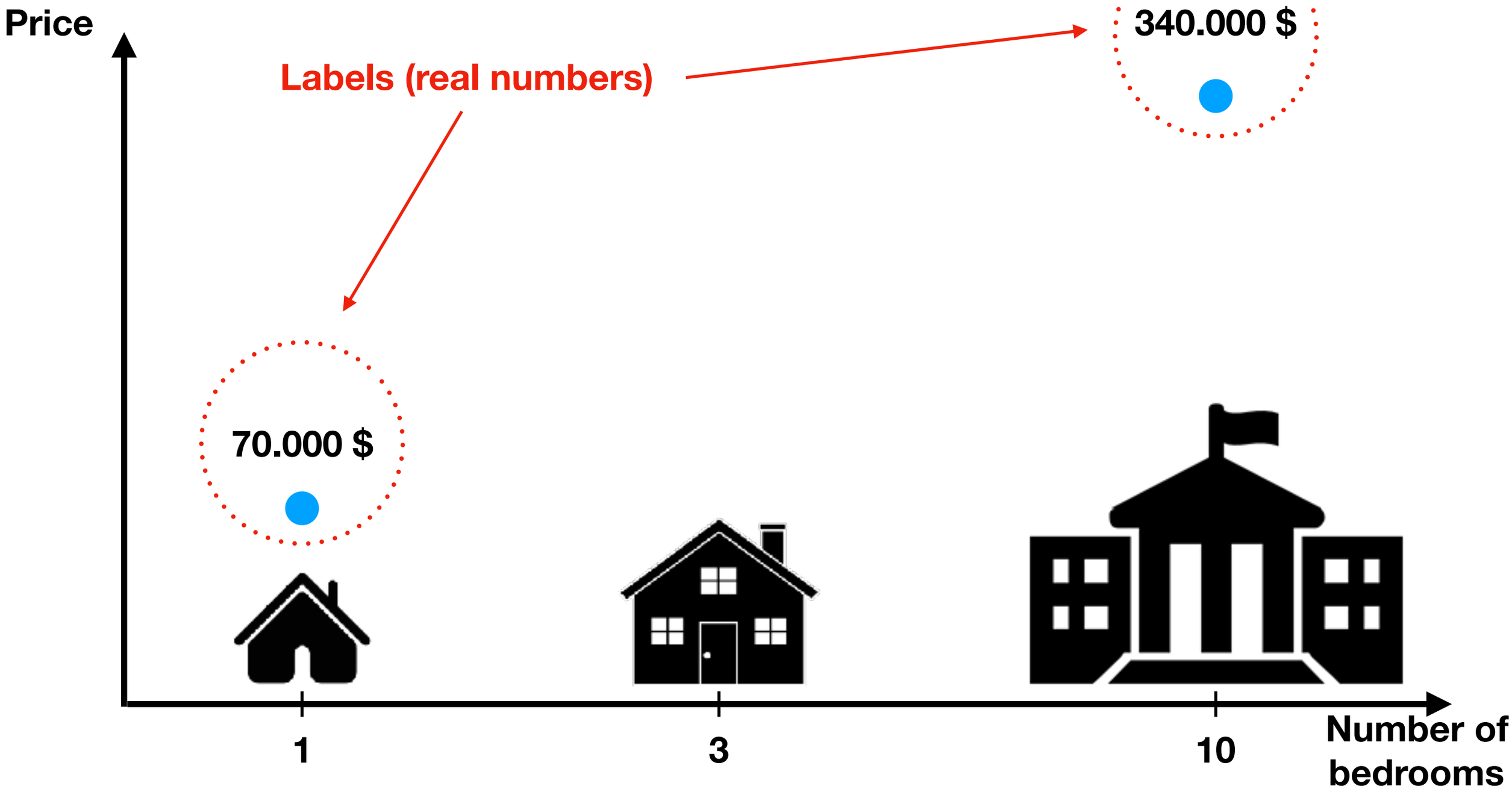
Regression



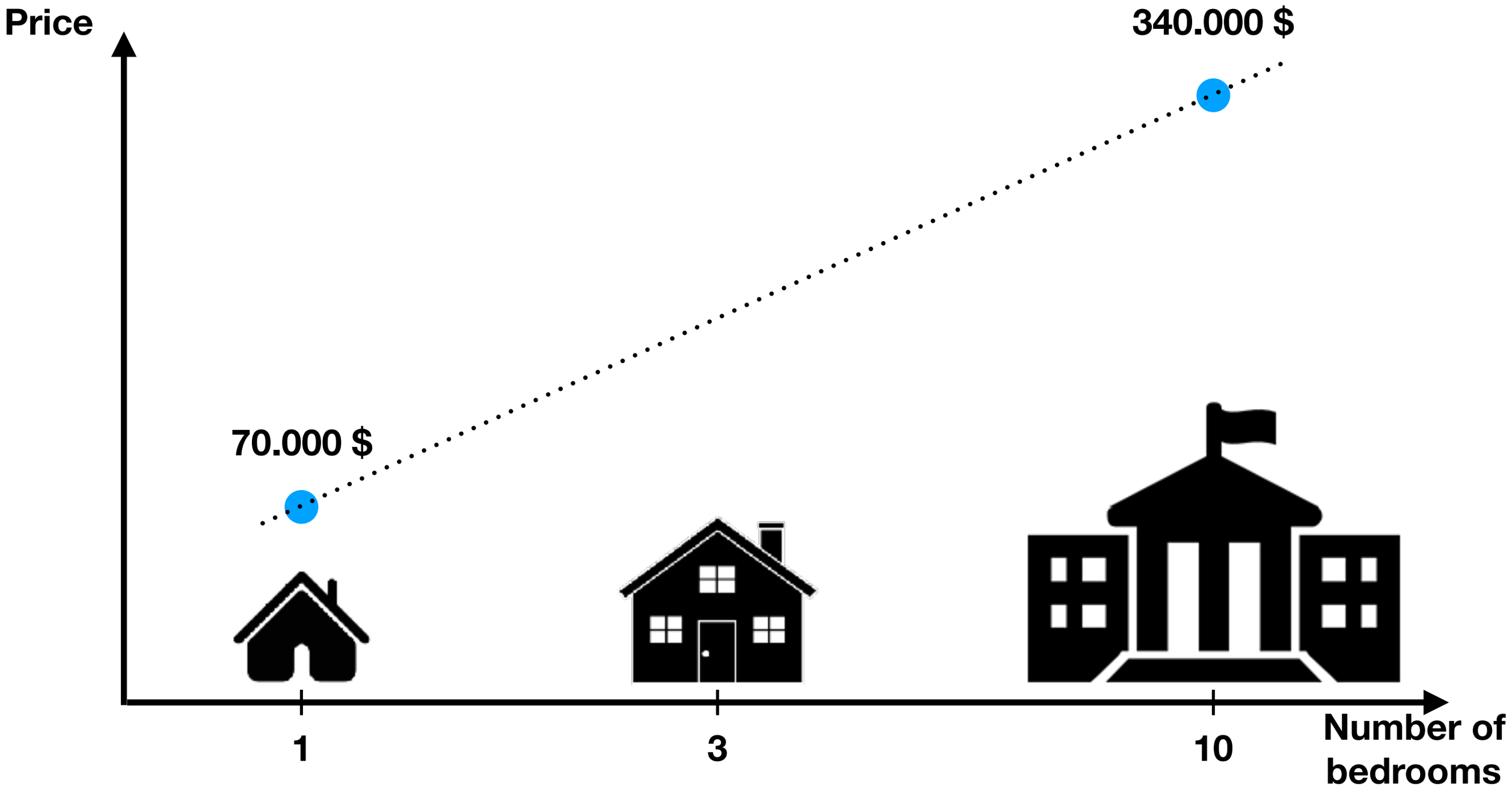
Regression



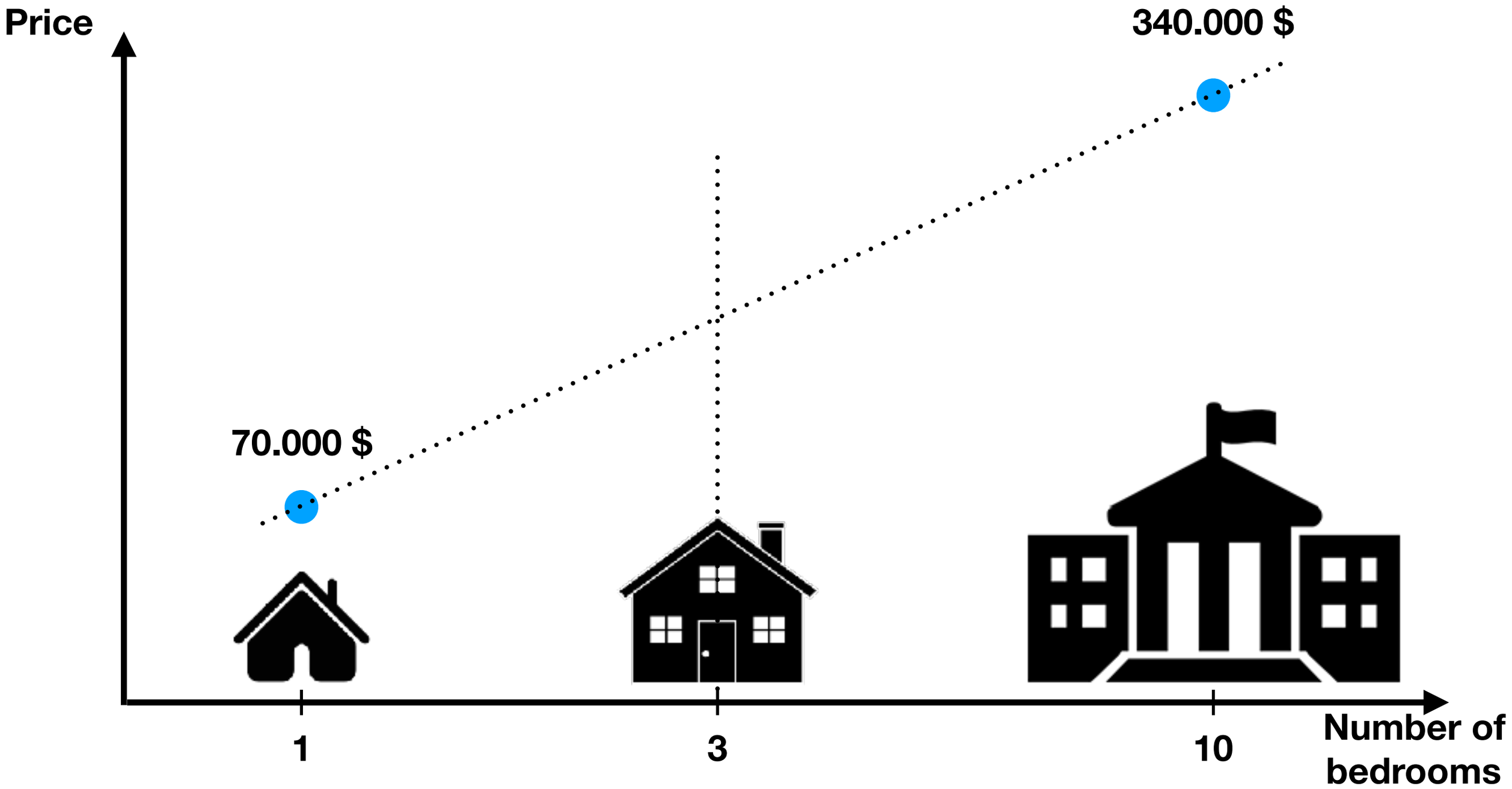
Regression


Price
Labels (real numbers)
340.000 $
70.000 $
1
3
10
Number of bedrooms

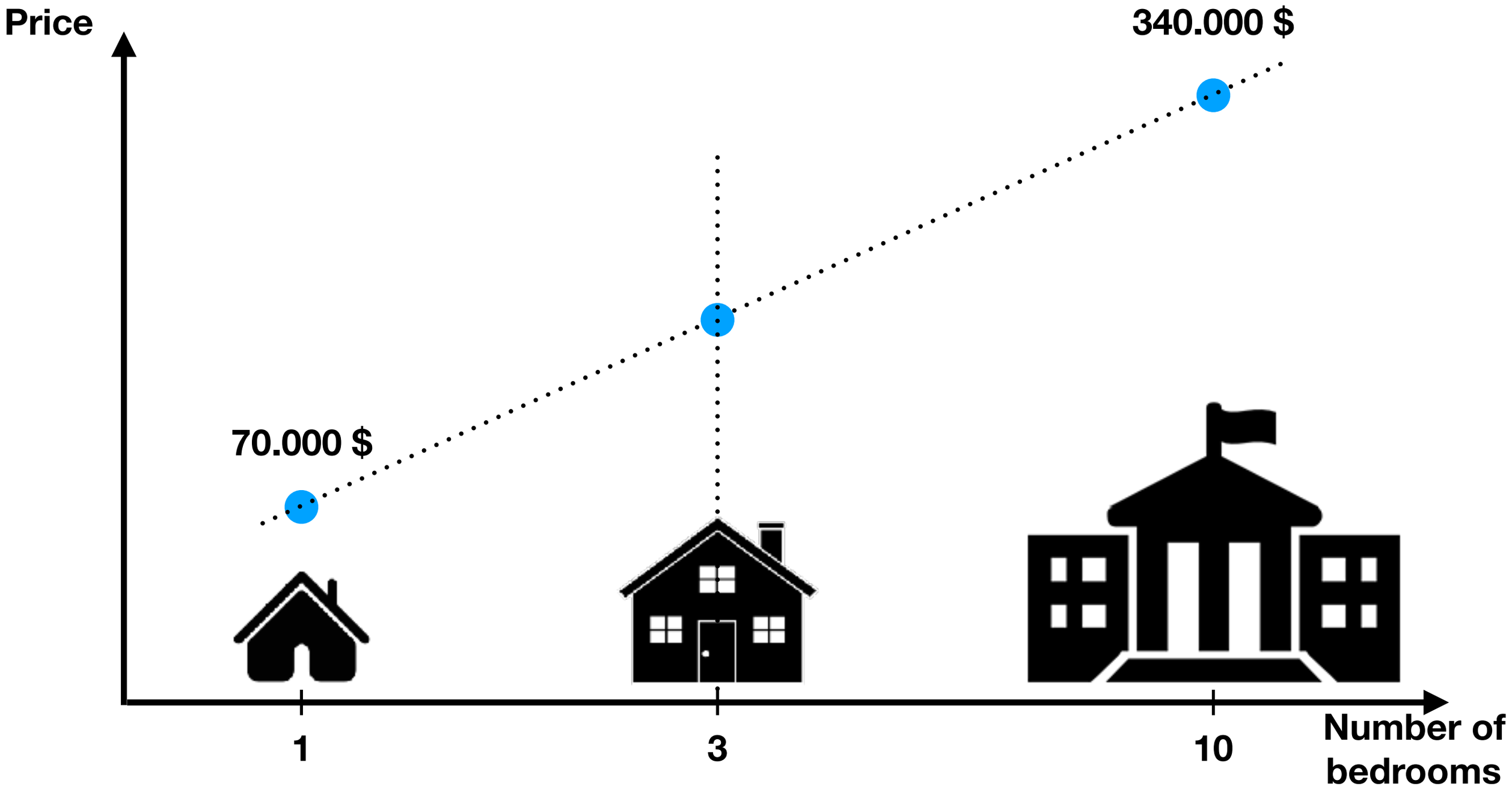
Regression



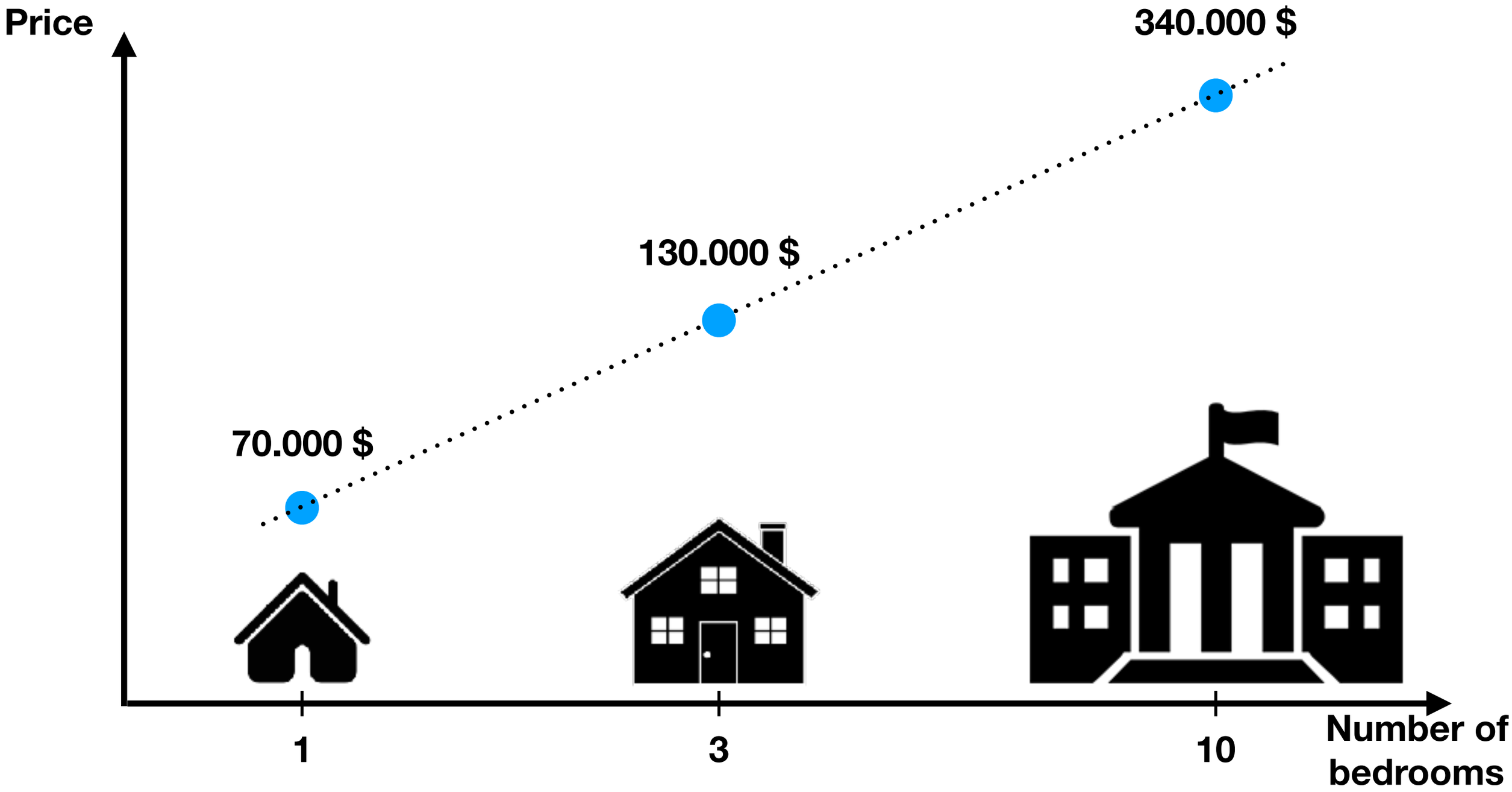
Regression



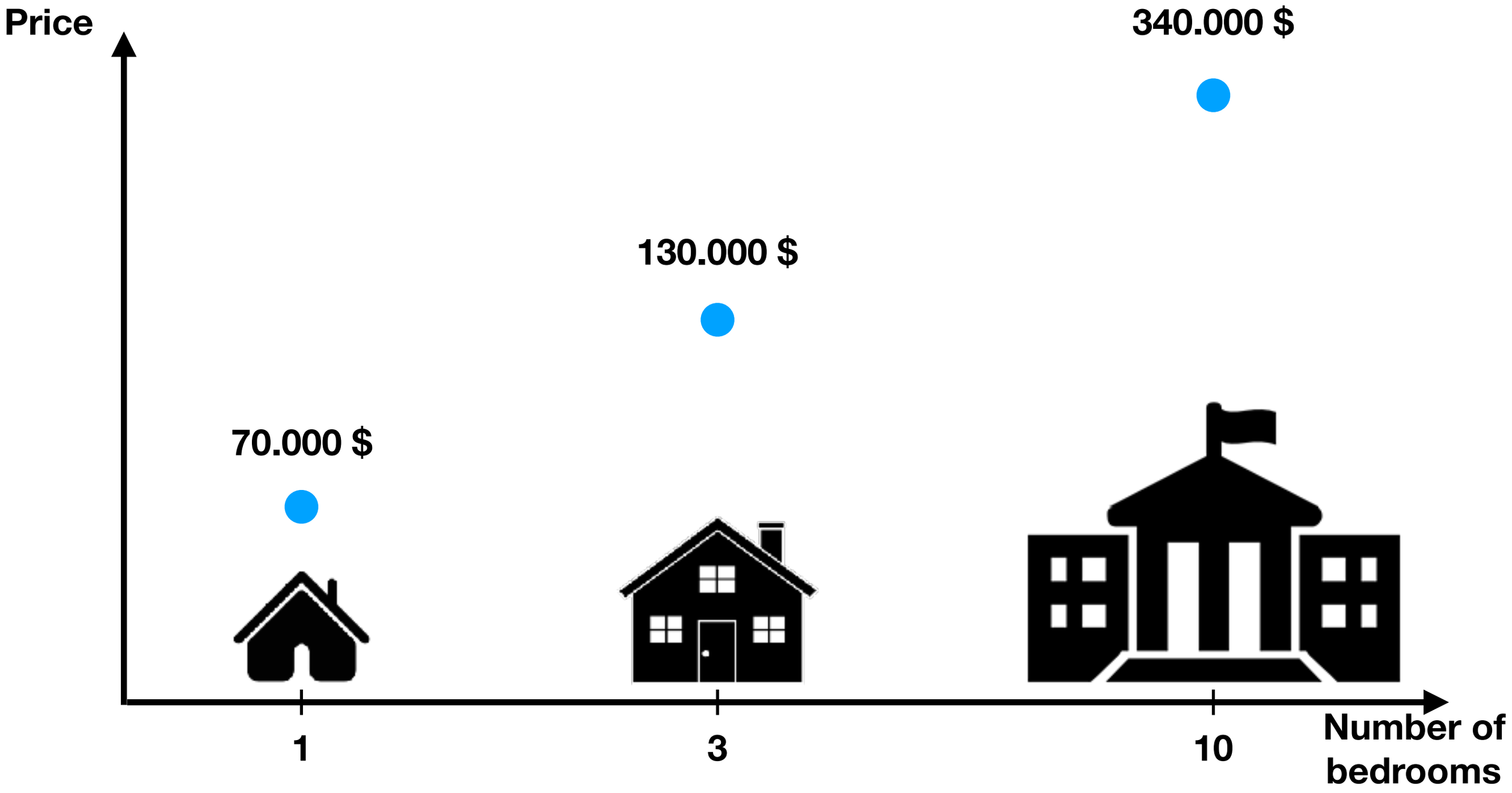
Regression



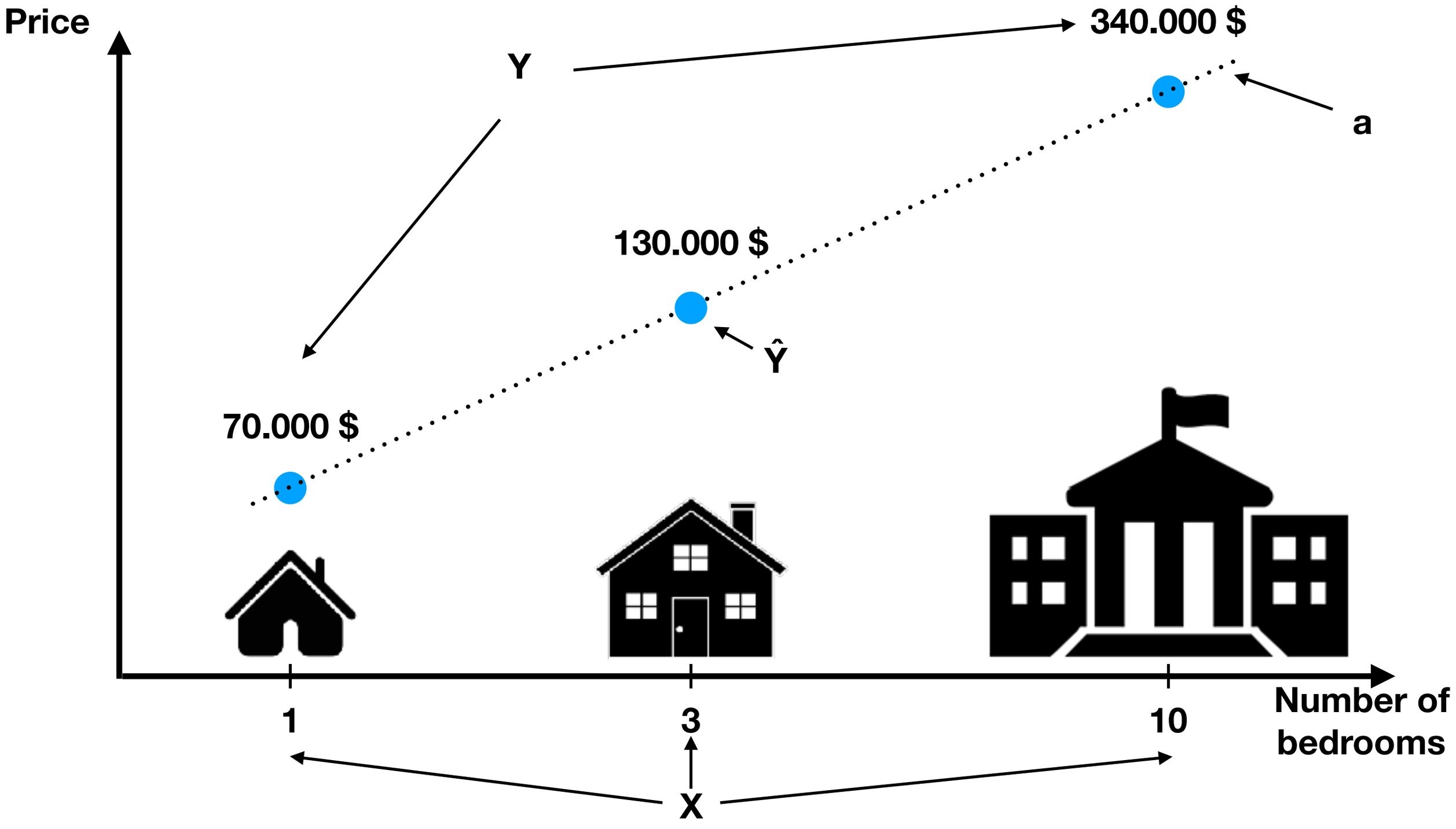
Regression



Regression

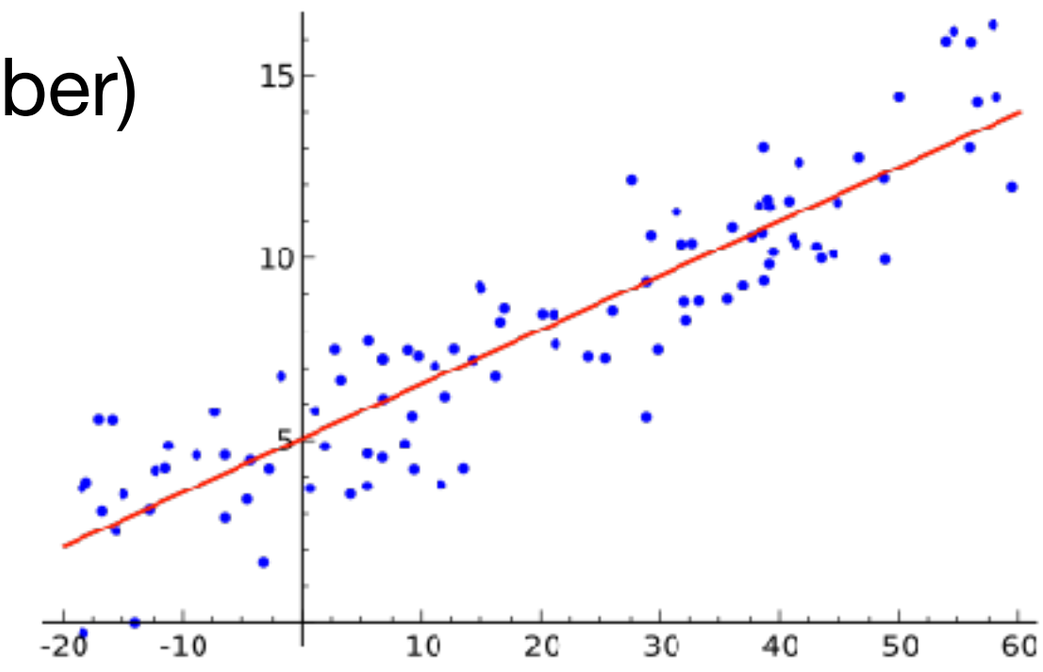


Regression



Regression

- If for **classification** we build a: X->Y
  when Y={0,1} or Y={class1, class2, class3, …}
- **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	**???**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

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	**???**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

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	**???**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

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	**???**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

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	**???**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

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

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

But what's better? ***Y1*** or ***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	**???**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

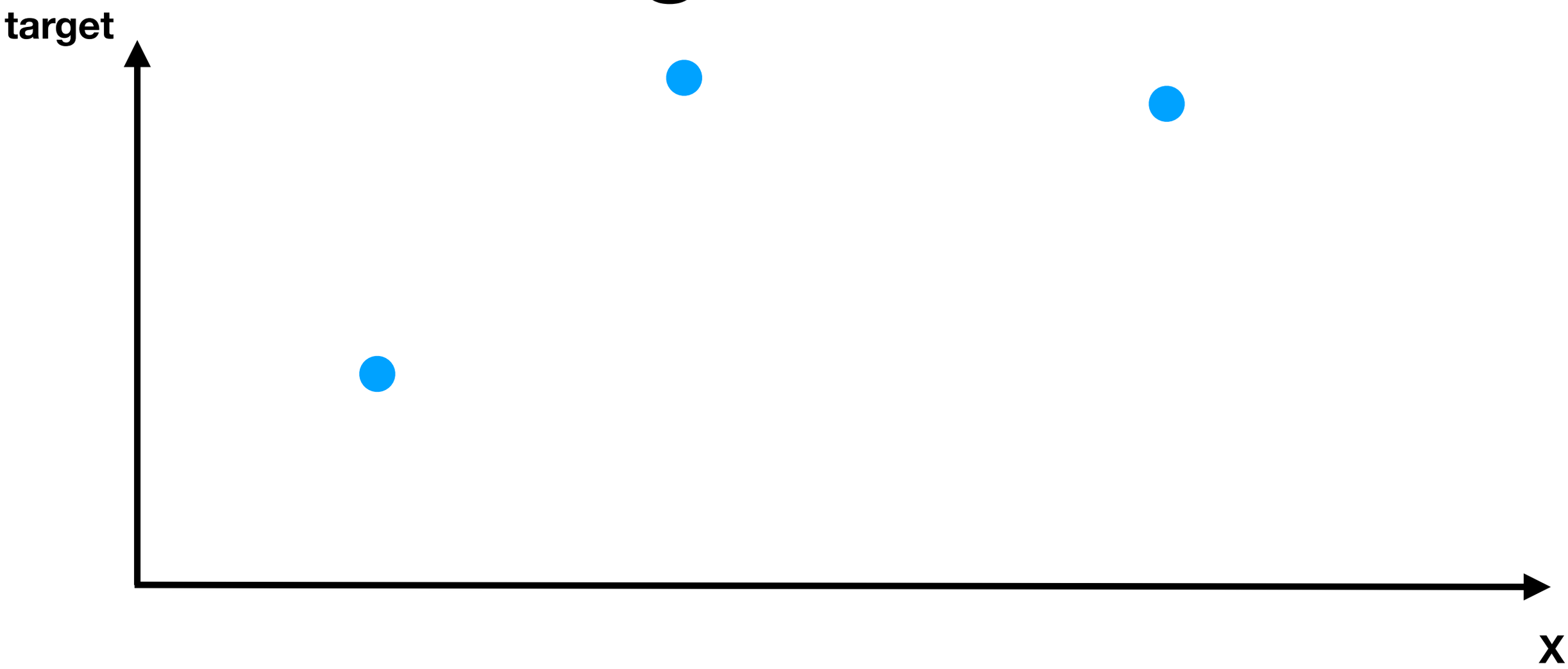
Need find: *a, b, c, d*

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

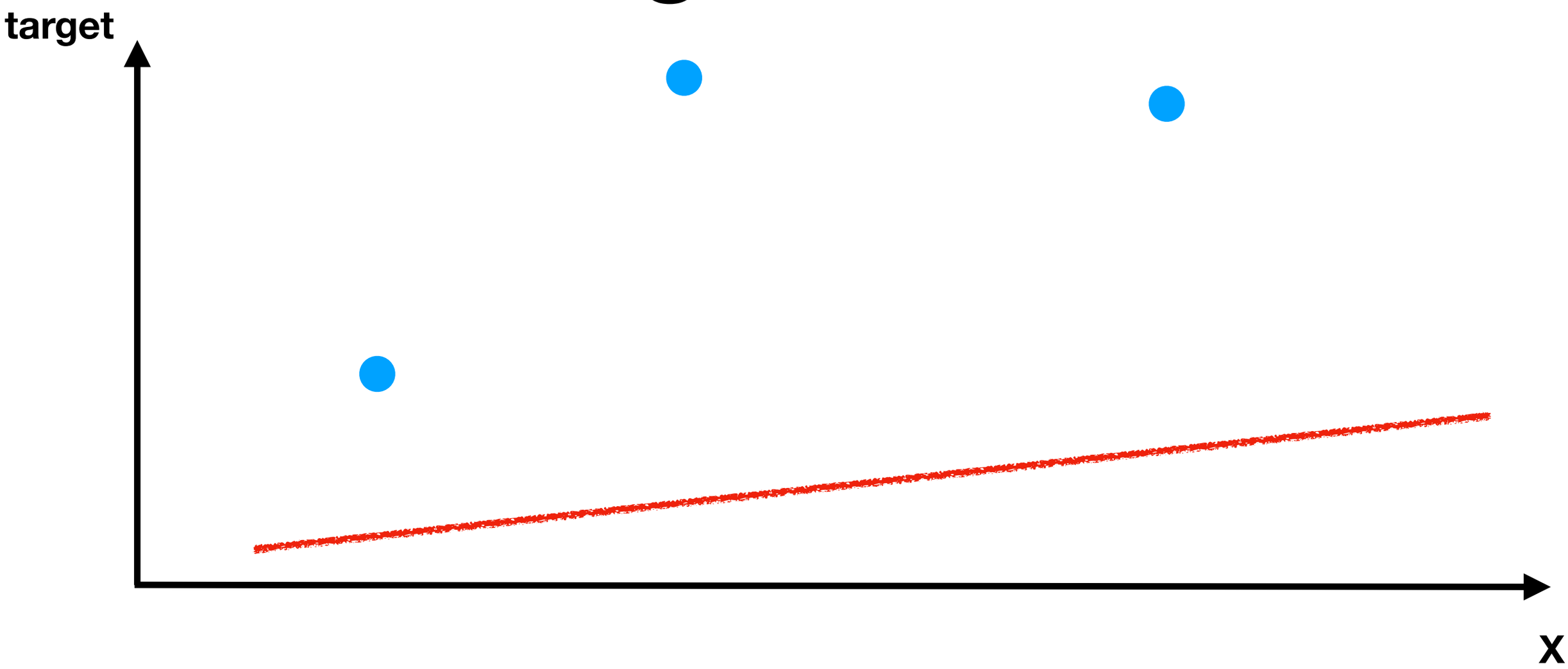
But what's better? ***Y1*** or ***Y2***

Need to define **Error**

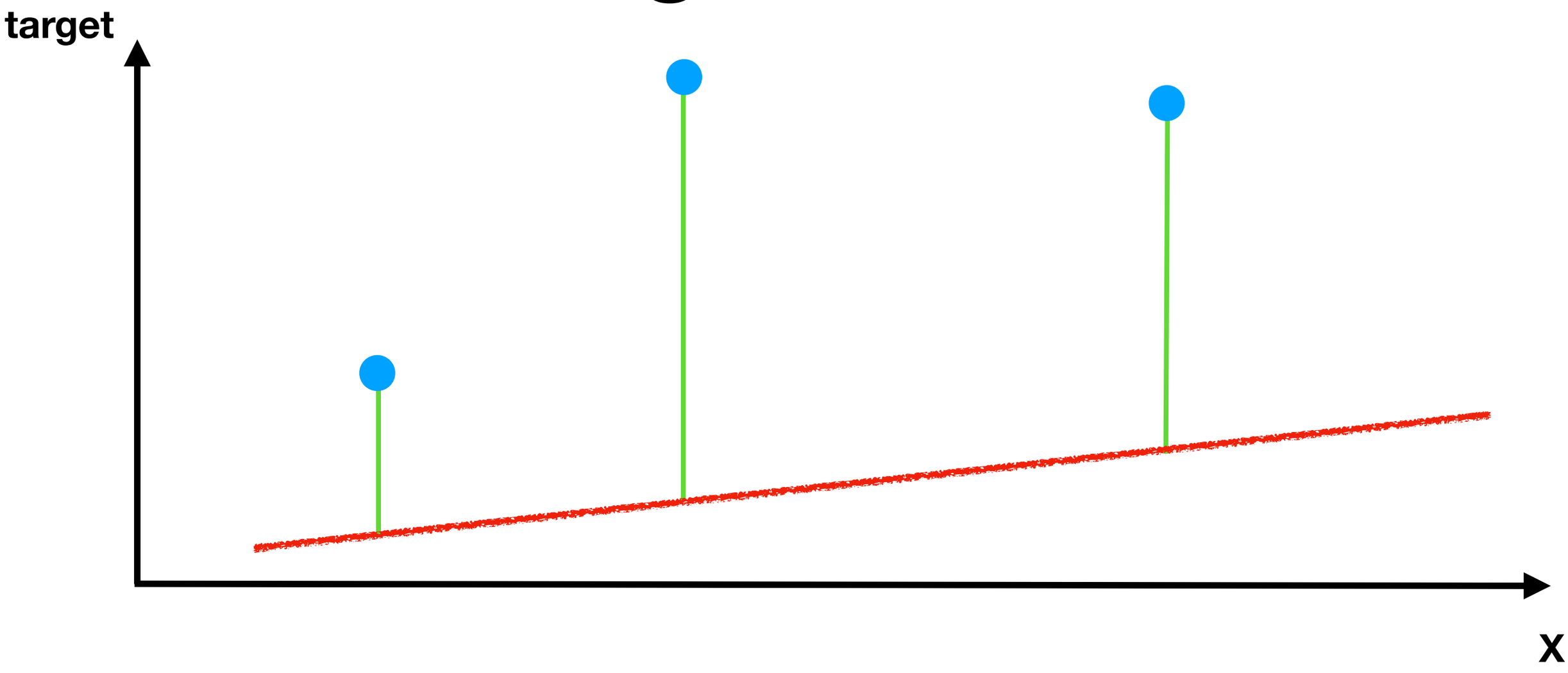
Good or bad regression



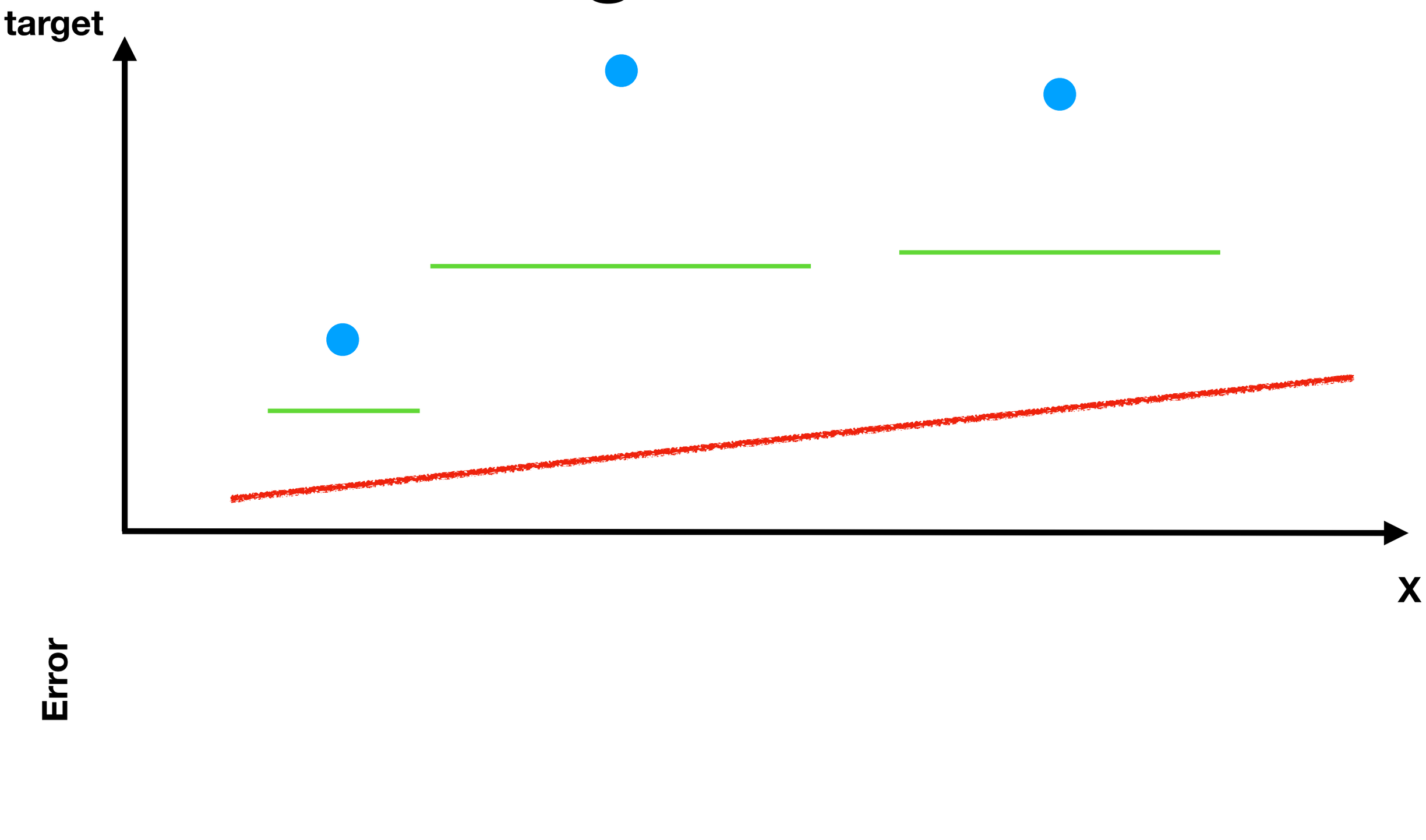
Good or bad regression



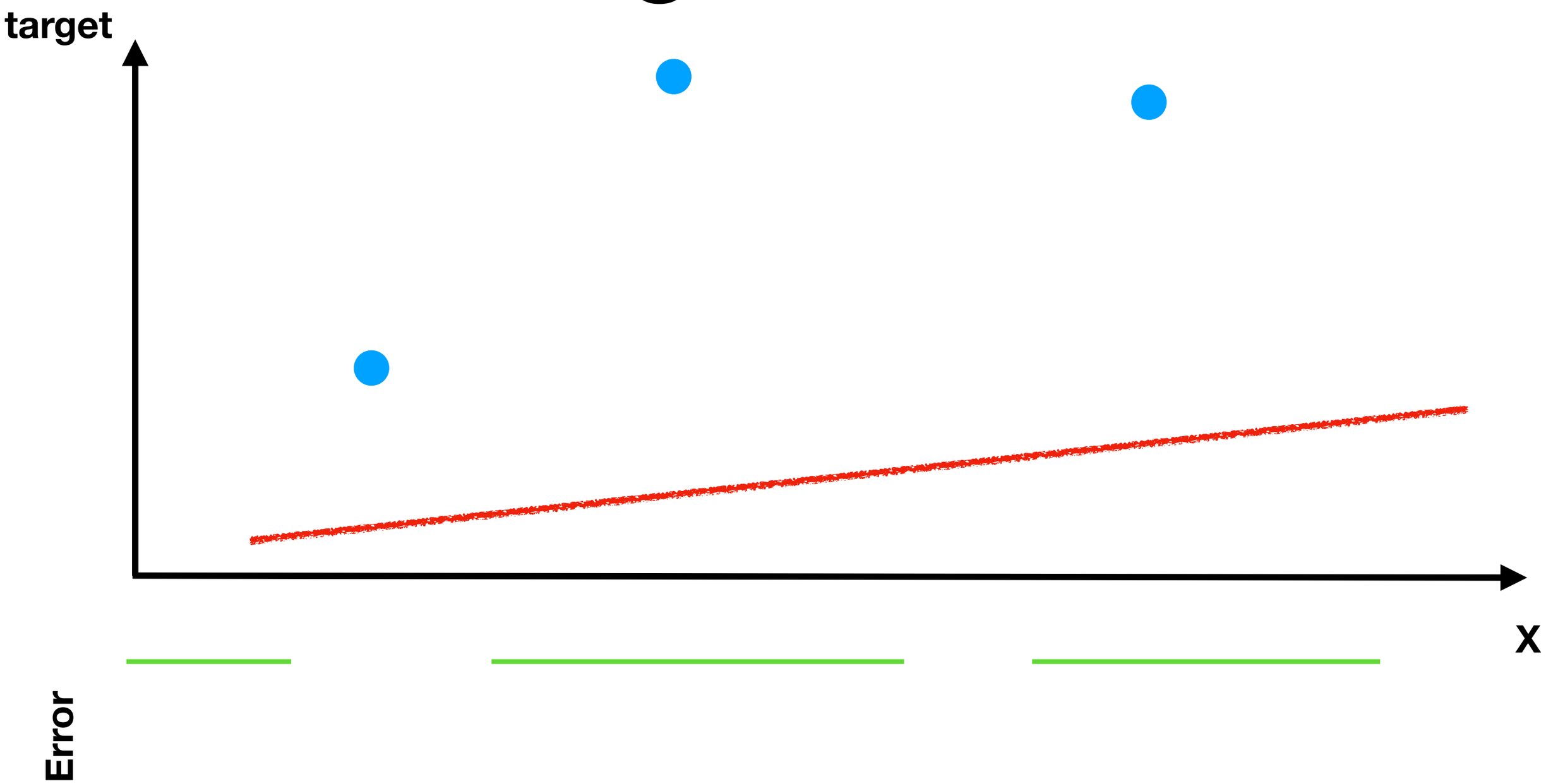
Good or bad regression



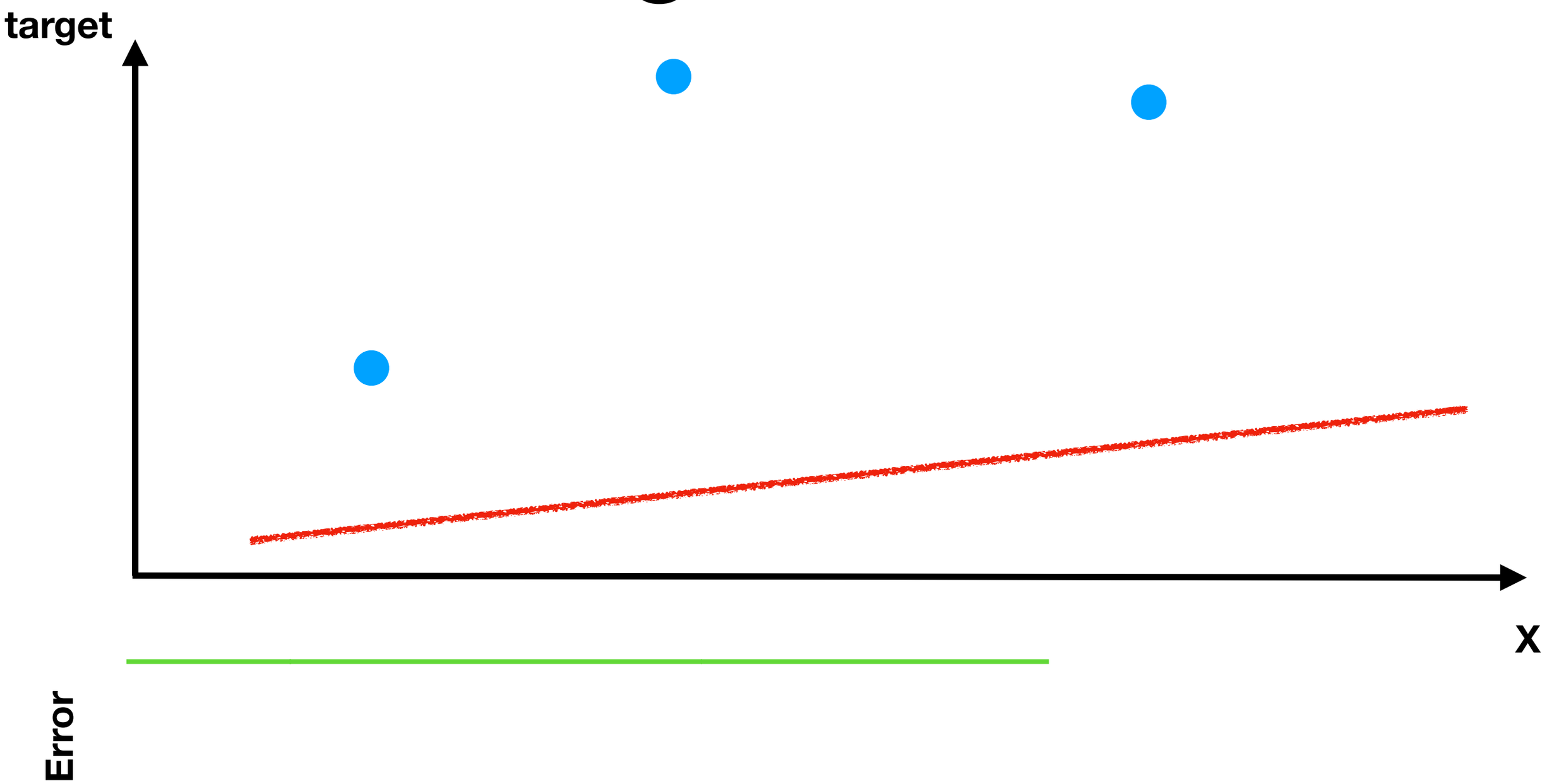
Good or bad regression



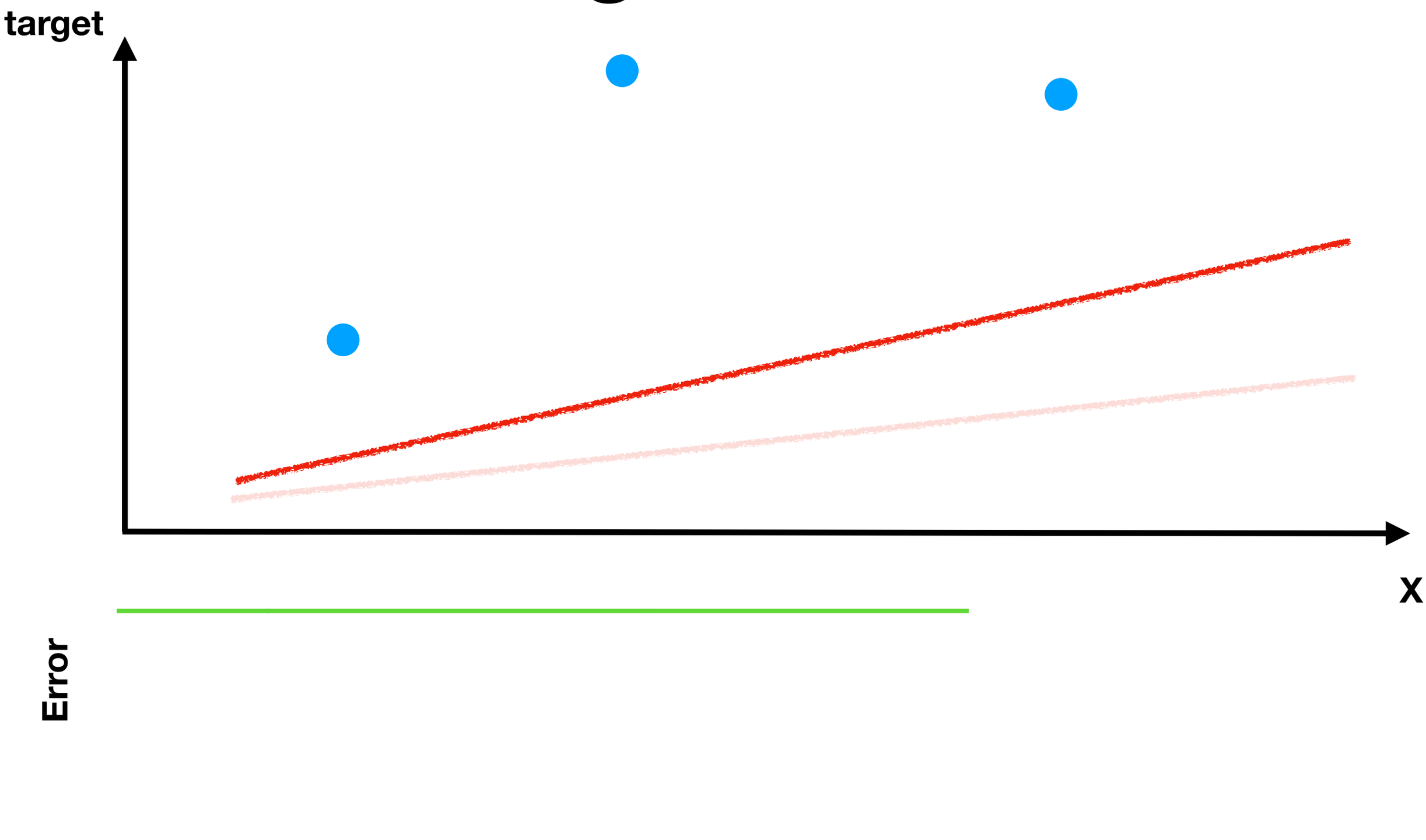
Good or bad regression



Good or bad regression

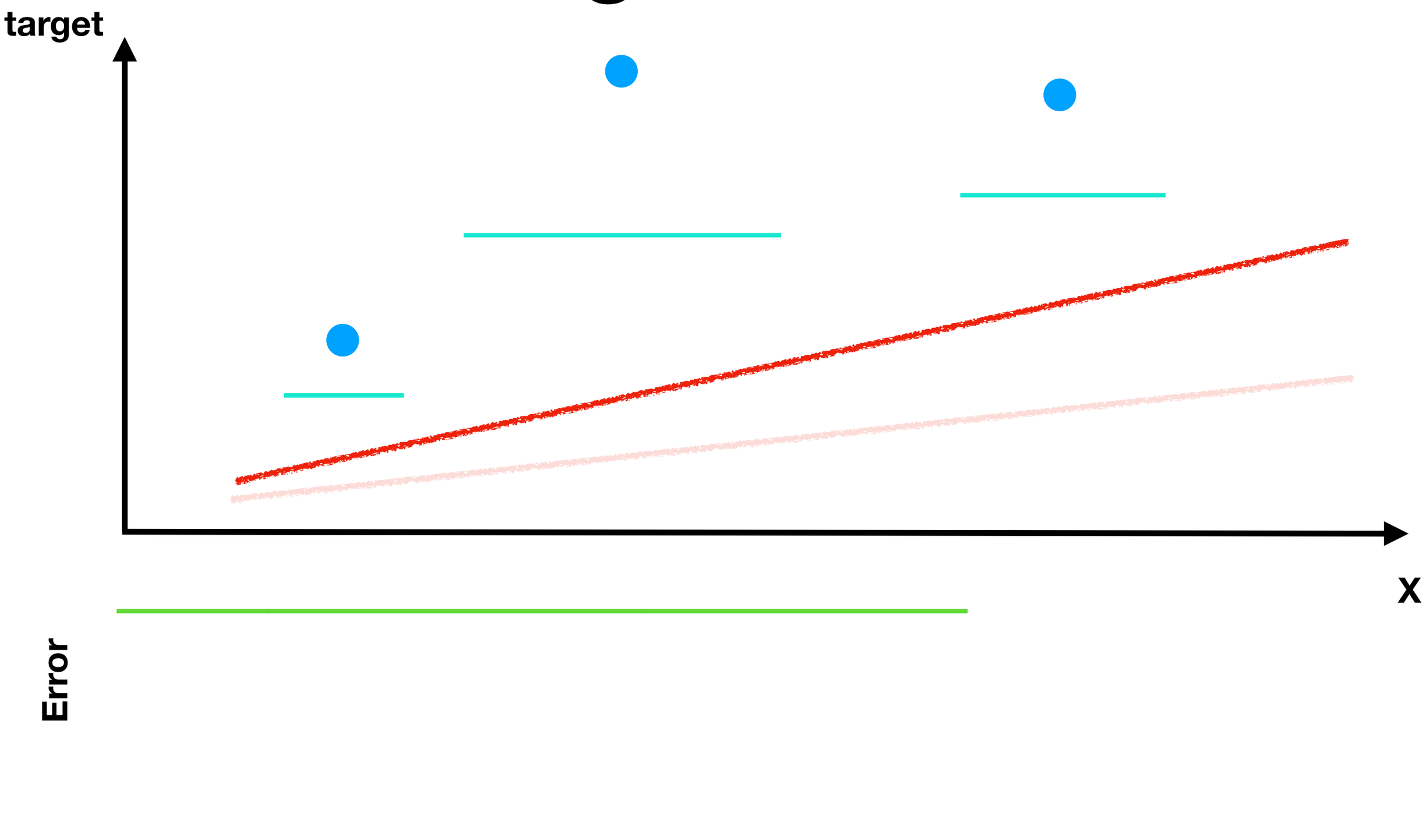


Good or bad regression

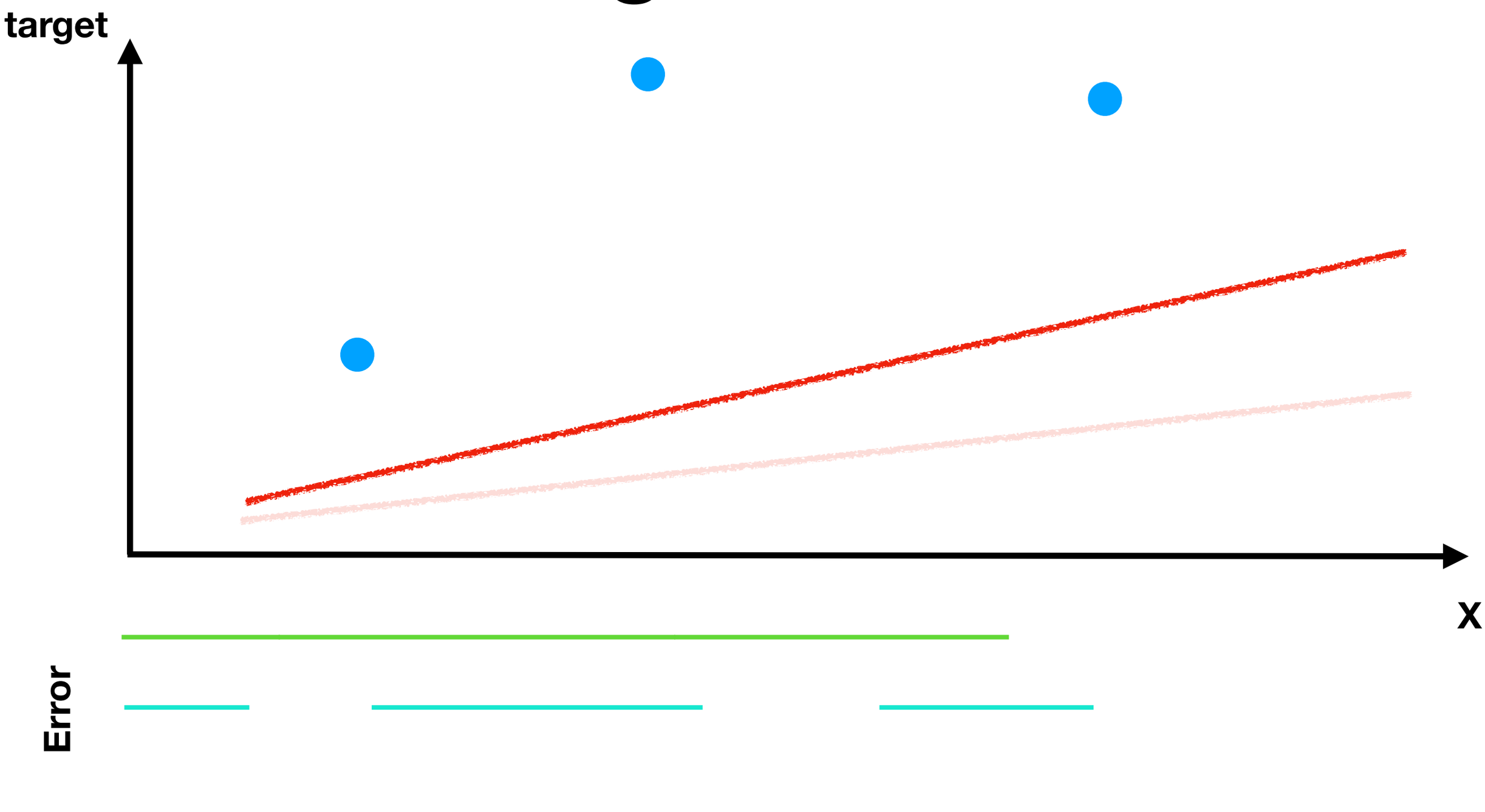


Good or bad regression

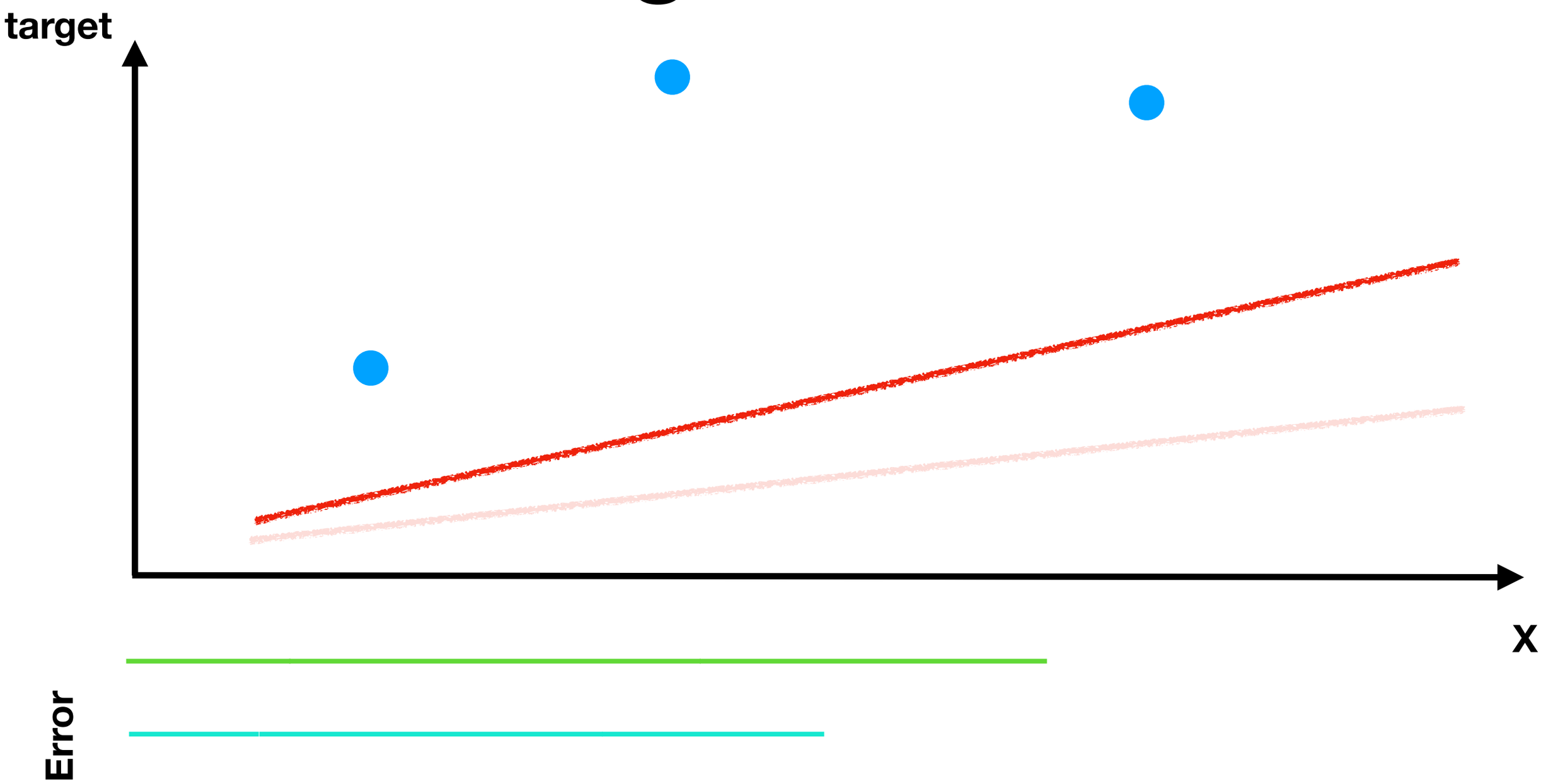
Good or bad regression



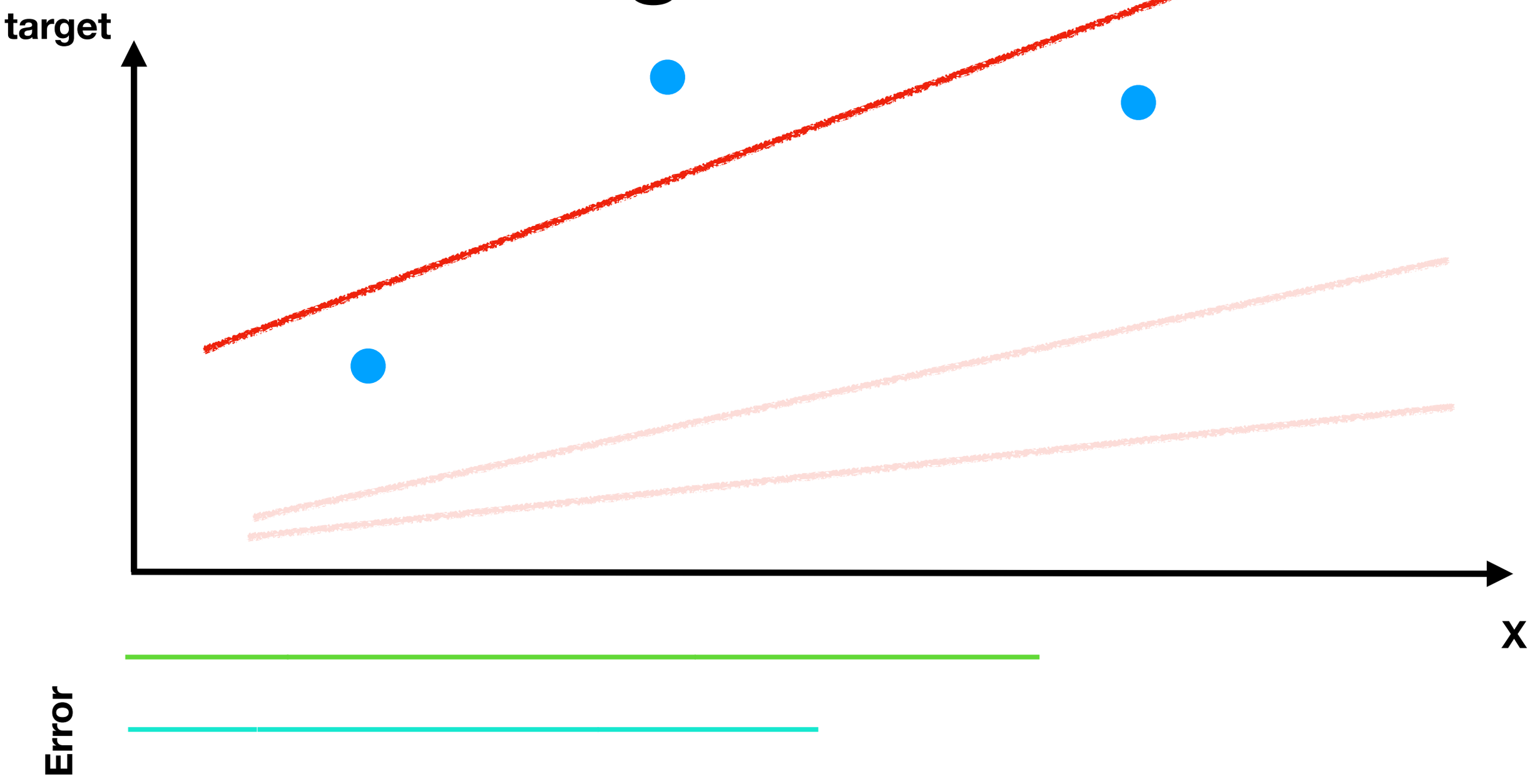
Good or bad regression



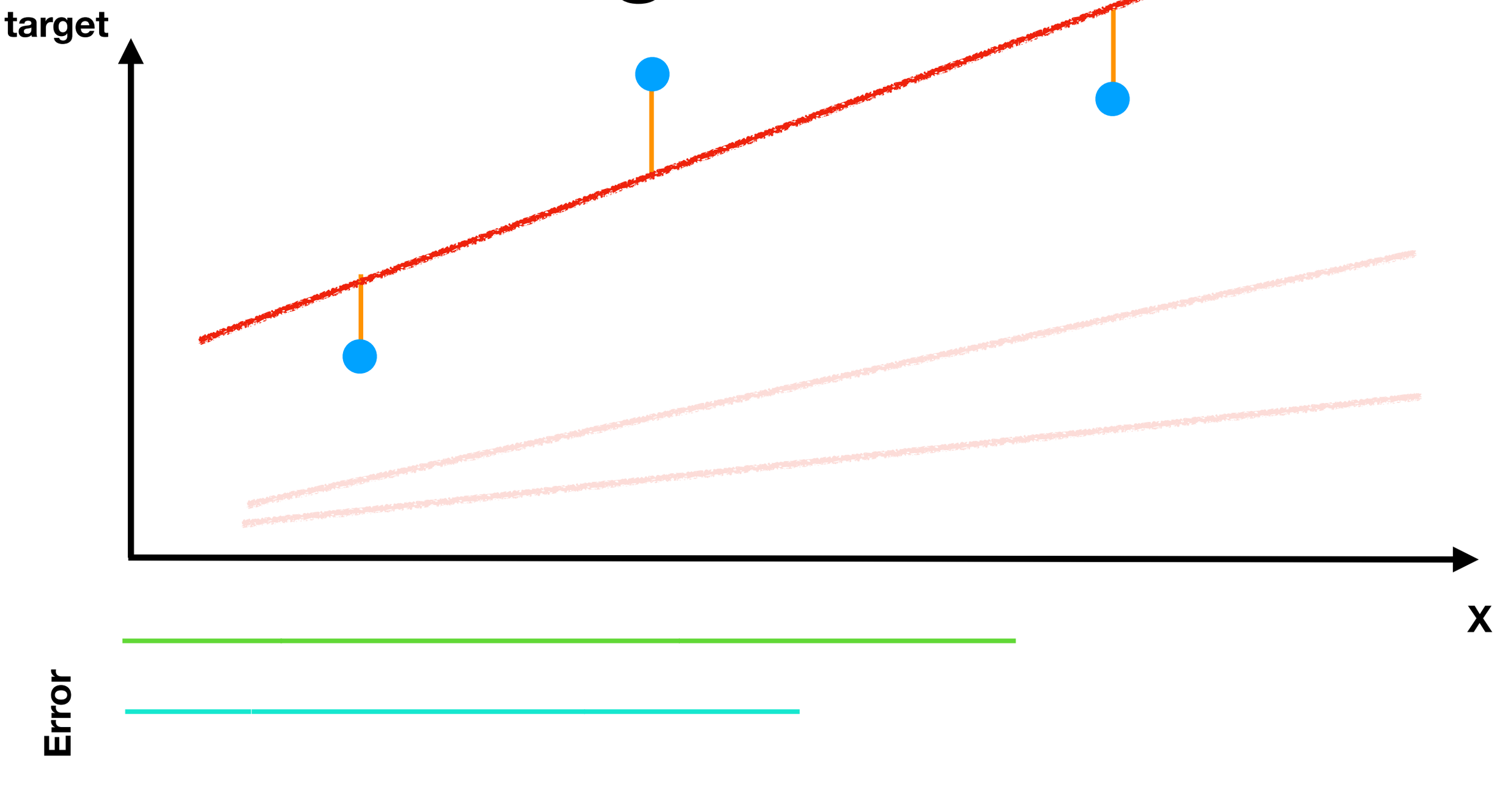
Good or bad regression



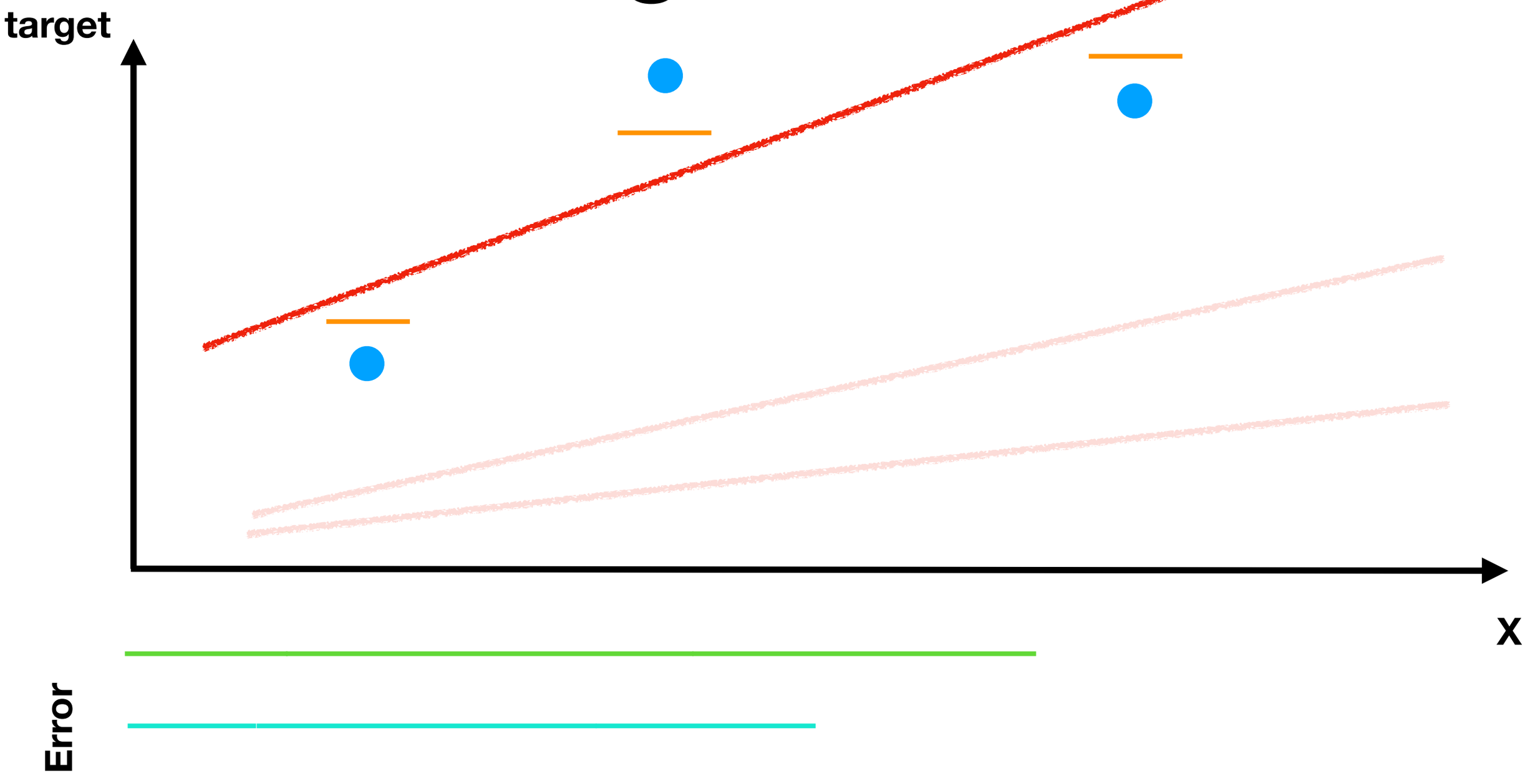
Good or bad regression



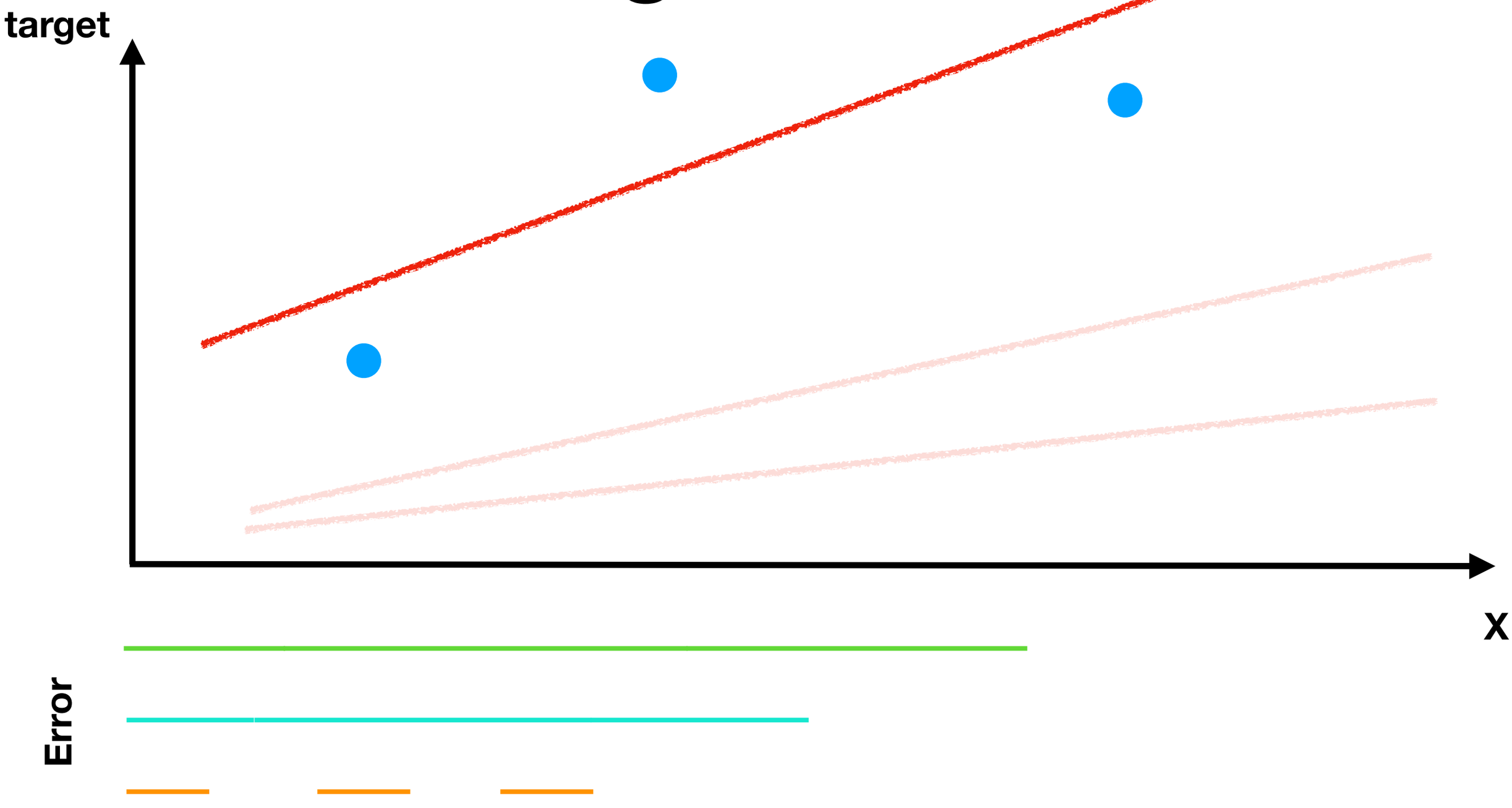
Good or bad regression



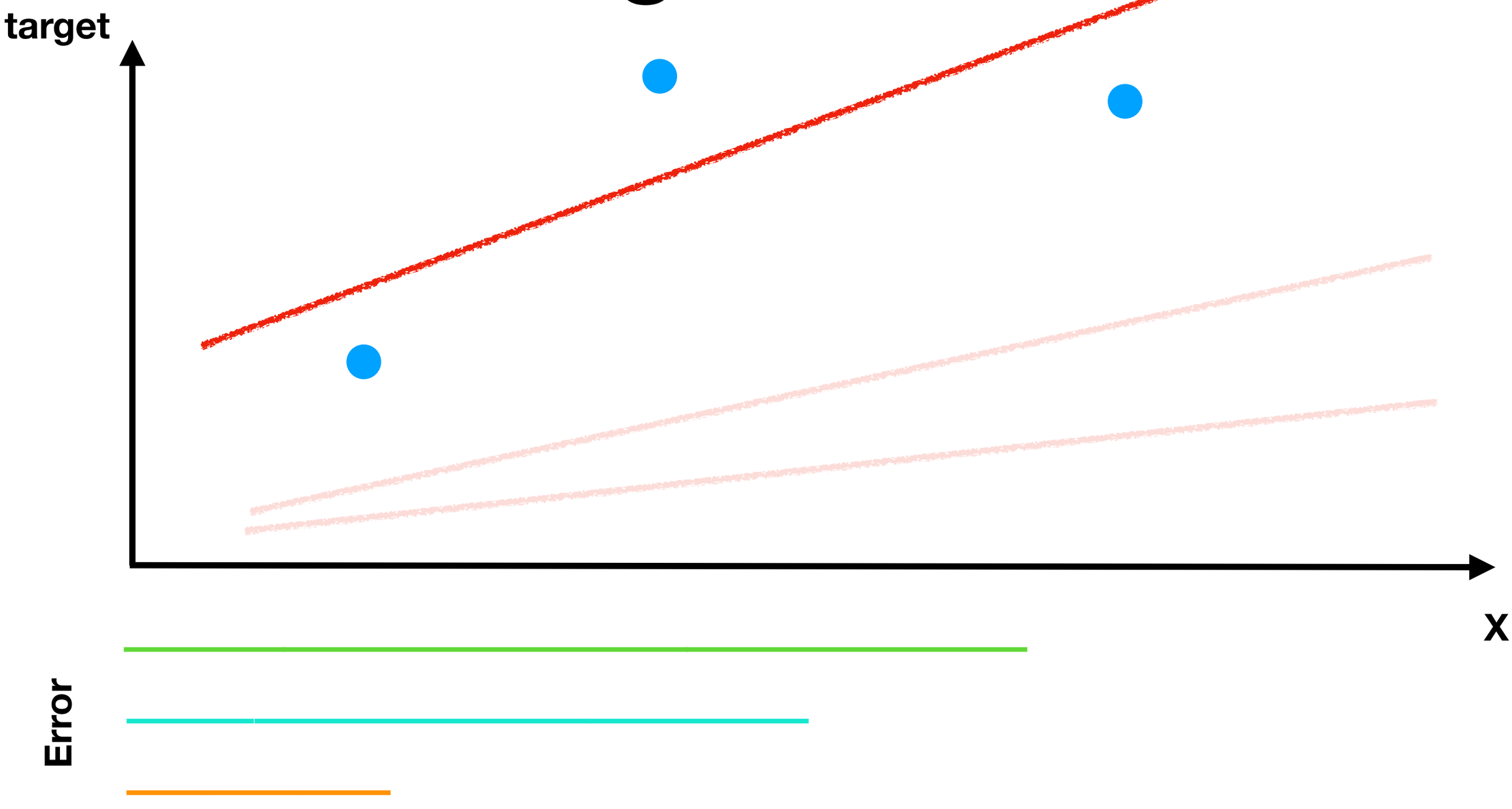
Good or bad regression



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	**???**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

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	**???**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

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	**???**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

Machine Learning Task:

find: *a, b, c, d*

and minimization of error

Regression metrics

Regression metrics

- MAE

Regression metrics

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

Regression metrics

- MAE $$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$
- MSE

Regression metrics

- MAE $$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$
- MSE $$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2$$

Regression metrics

- MAE $$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$
- MSE $$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2$$
- RMSE

Regression metrics

- MAE $$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$
- MSE $$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2$$
- RMSE $$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Regression metrics

- MAE $$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$
- MSE $$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2$$
- RMSE $$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$
- RMSLE

Regression metrics

- MAE $$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$
- MSE $$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2$$
- RMSE $$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$
- RMSLE $$\mathrm{RMSLE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(\log(y_j + 1) - \log(\hat{y}_j + 1))^2}$$

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	**???**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

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	**???**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

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

$$\hat{Y} = a_{best}*X1 + b_{best}*X2 + c_{best}*X3 + d_{best}$$

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	**???**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a^*X1 + b^*X2 + c^*X3 + d$$

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

$$\hat{Y} = a_{best}^*X1 + b_{best}^*X2 + c_{best}^*X3 + d_{best}$$

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

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	**???**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

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

$$\hat{Y} = a_{best}*X1 + b_{best}*X2 + c_{best}*X3 + d_{best}$$

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

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	**???**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a^*X1 + b^*X2 + c^*X3 + d$$

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

$$\hat{Y} = a_{best}^*X1 + b_{best}^*X2 + c_{best}^*X3 + d_{best}$$

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

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**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a^*X1 + b^*X2 + c^*X3 + d$$

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

$$\hat{Y} = a_{best}^*X1 + b_{best}^*X2 + c_{best}^*X3 + d_{best}$$

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

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**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

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

$$\hat{Y} = a_{best}*X1 + b_{best}*X2 + c_{best}*X3 + d_{best}$$

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

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**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

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**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

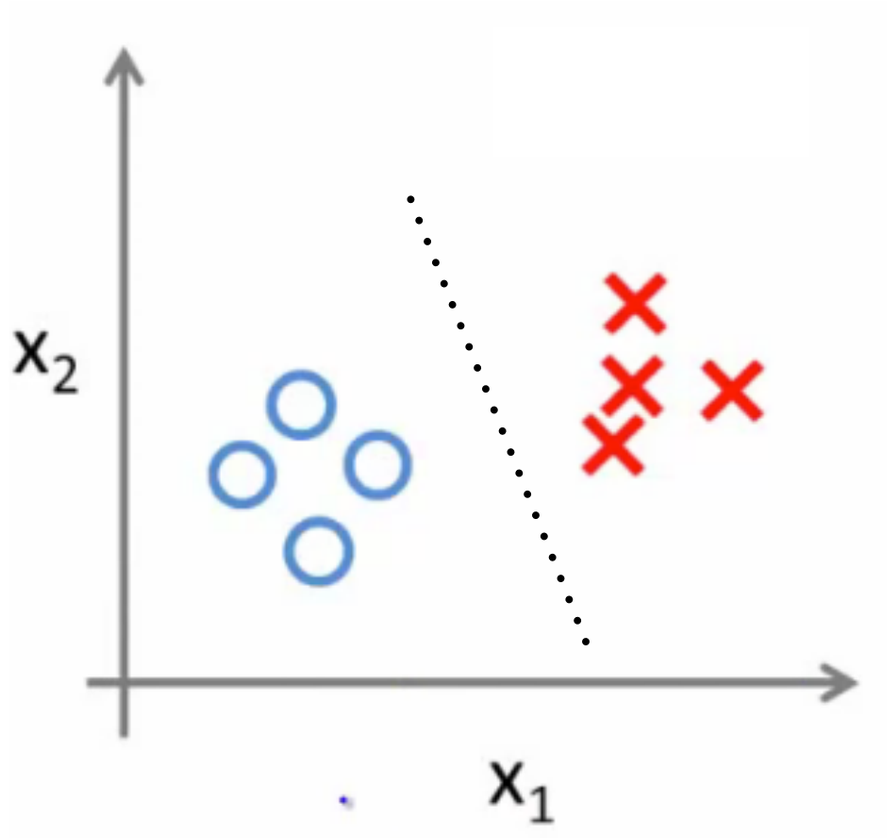
All values of features are real numbers

Classification

- $\hat{Y} = a*X1 + b*X2 + c*X3 + d$ - *regression example*

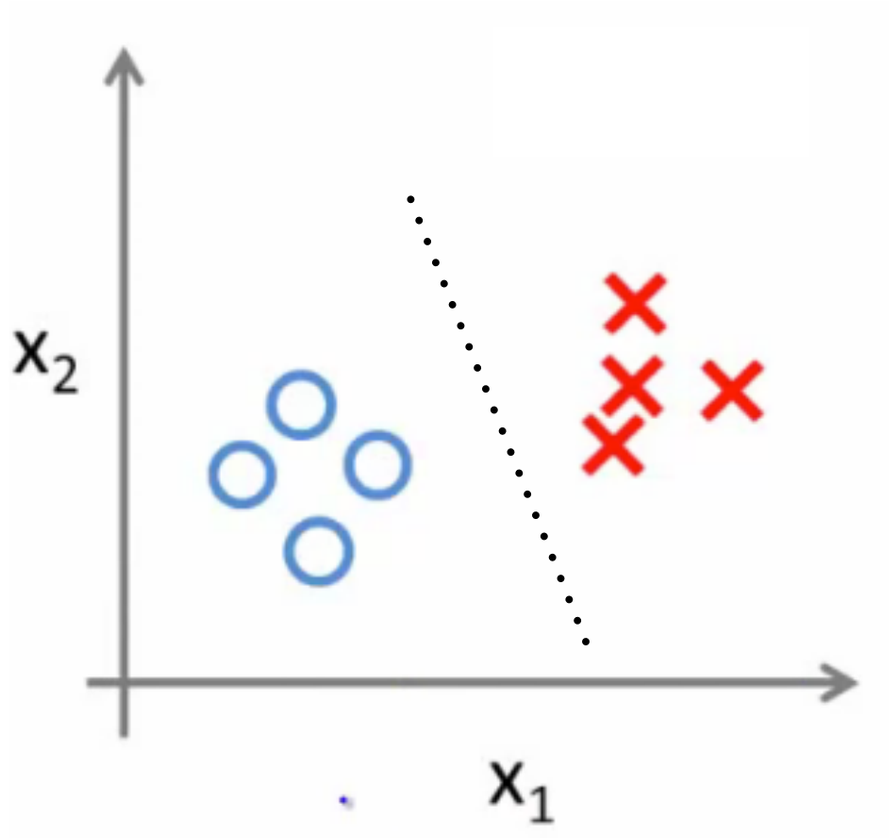
Classification

- $\hat{Y} = a*X1 + b*X2 + c*X3 + d$ - *regression example*



Classification

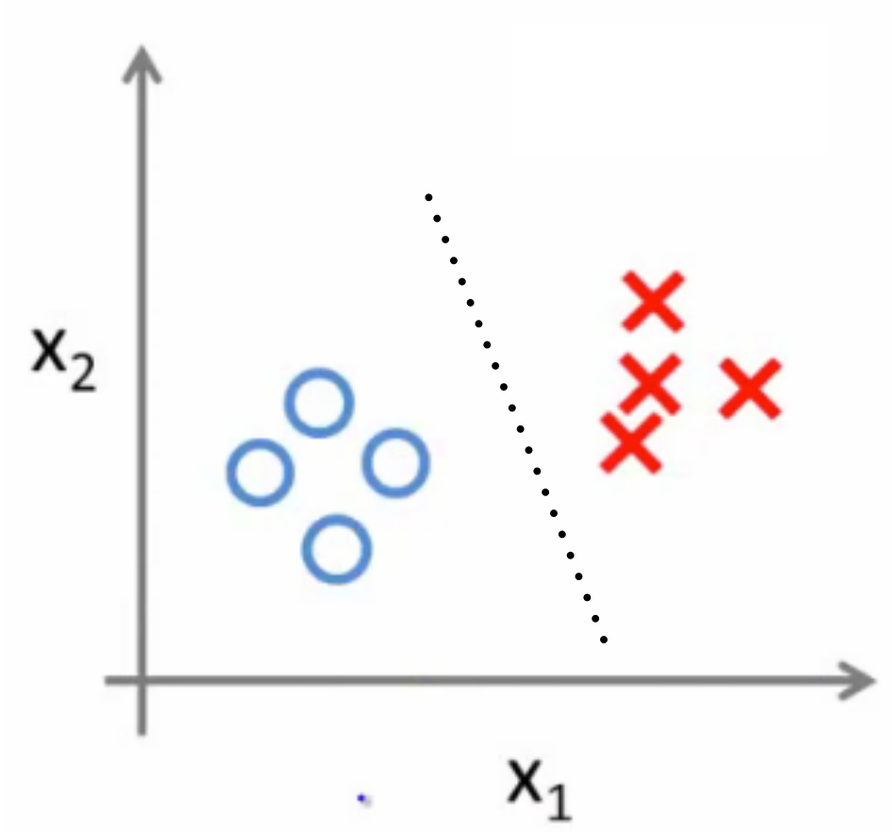
- $\hat{Y} = a*X1 + b*X2 + c*X3 + d$ - *regression example*
- $\hat{Y} = f(X)$



Classification

- $\hat{Y} = a*X1 + b*X2 + c*X3 + d$ - *regression example*
- $\hat{Y} = f(X)$

$\hat{Y} = a*X1 + b*X2 + c$



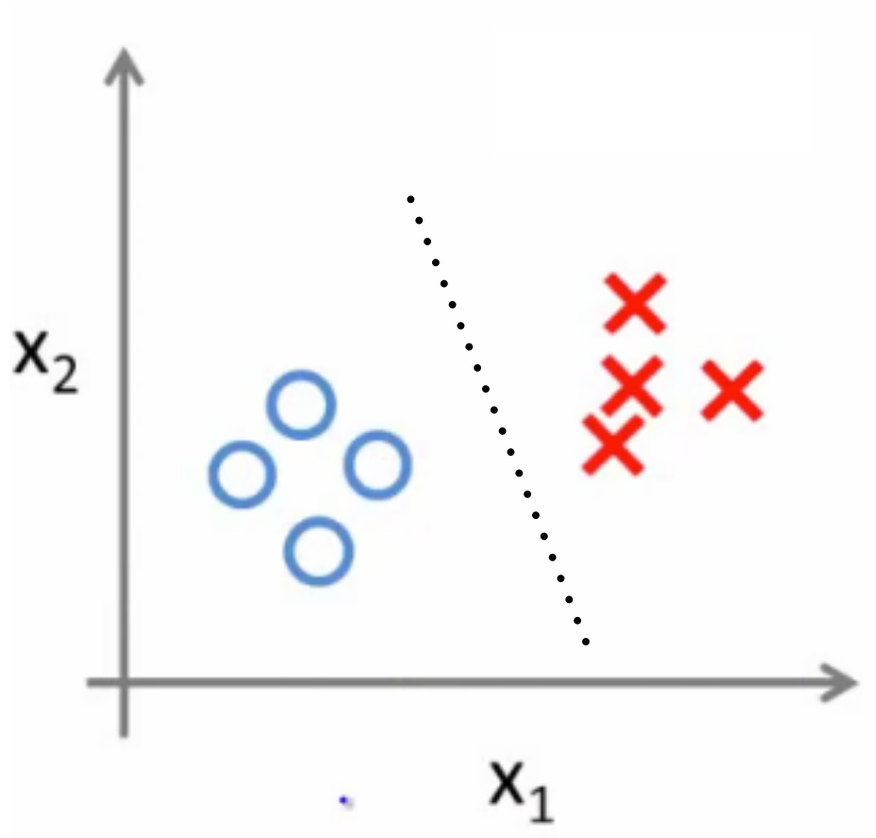
Classification

- $\hat{Y} = a*X1 + b*X2 + c*X3 + d$ - *regression example*
- $\hat{Y} = f(X)$

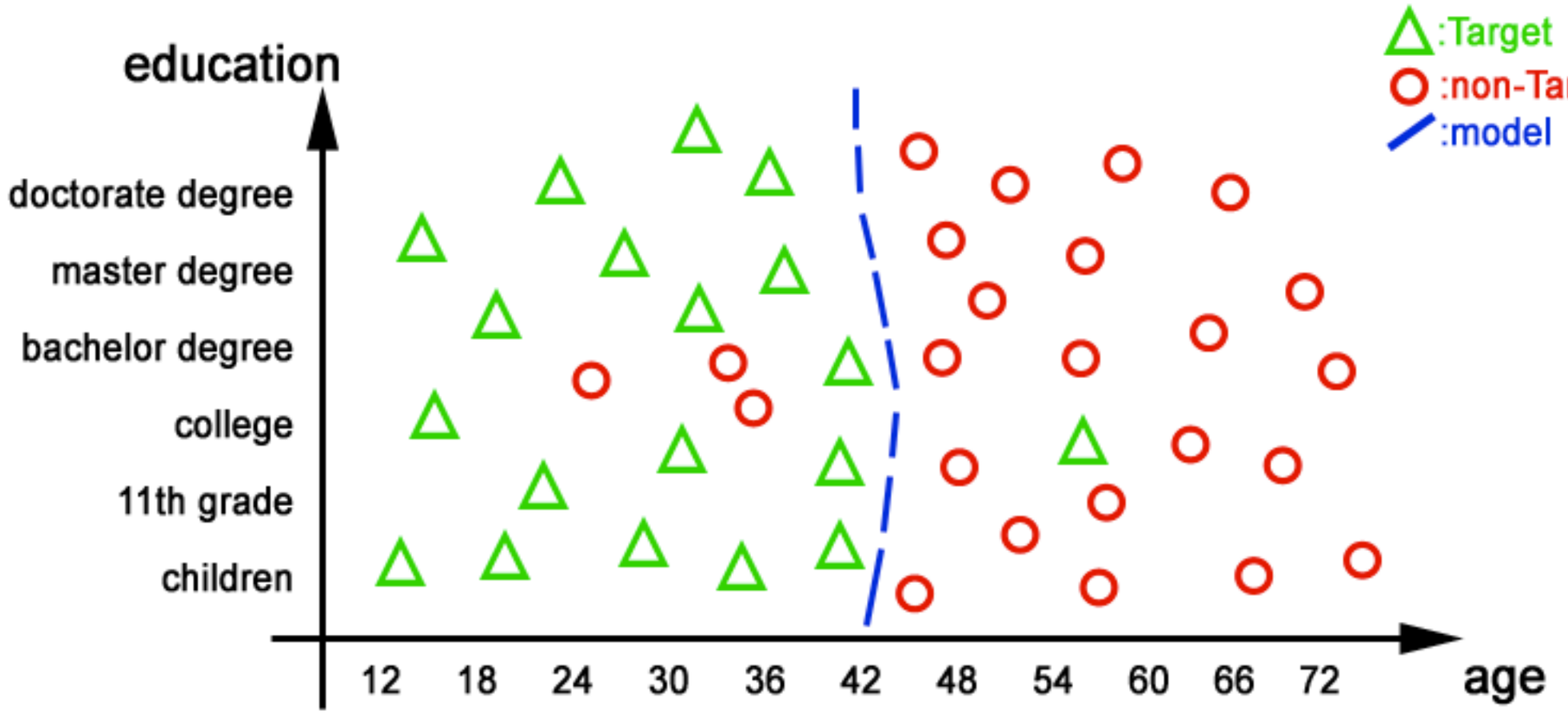
$\hat{Y} = a*X1 + b*X2 + c$

$\hat{Y} = f(X) > 0$ ×

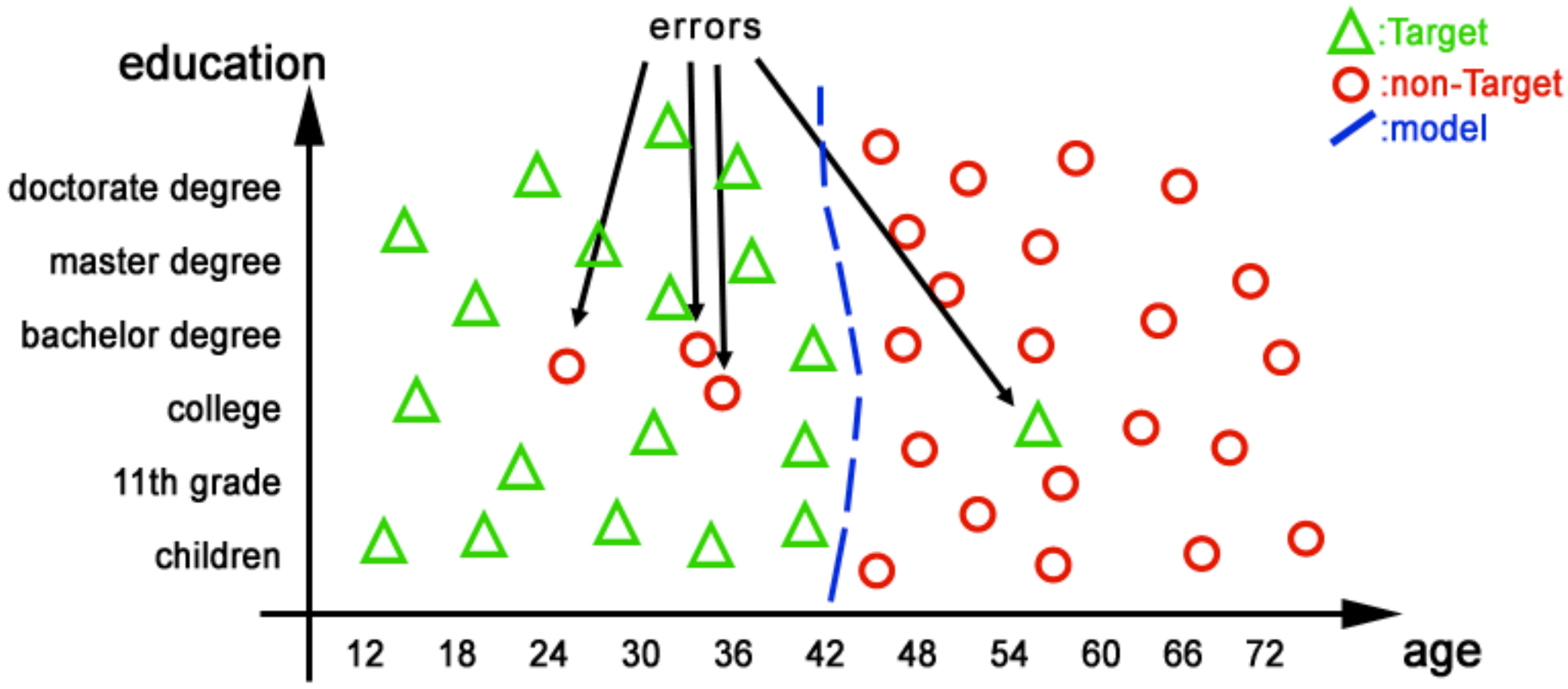
$\hat{Y} = f(X) < 0$ ○



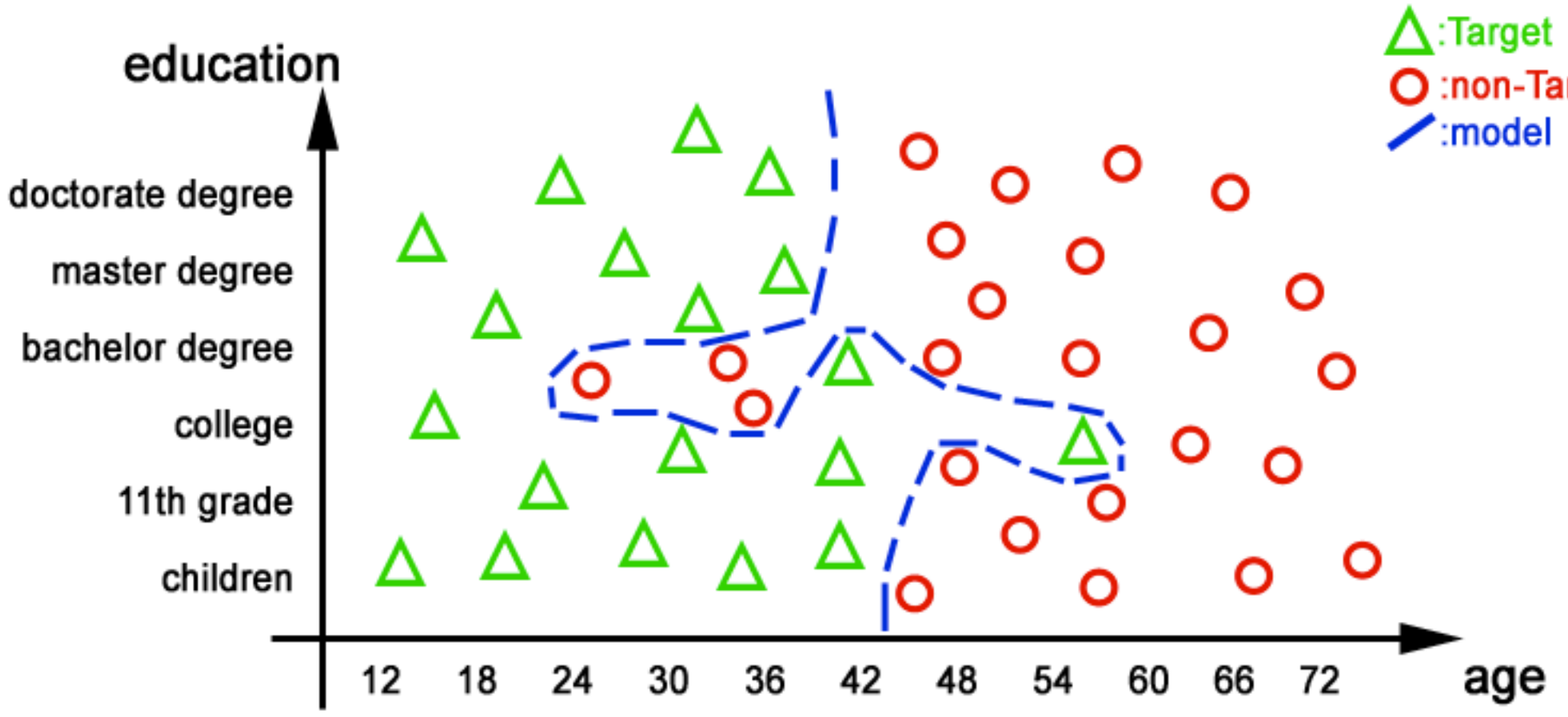
Model with errors



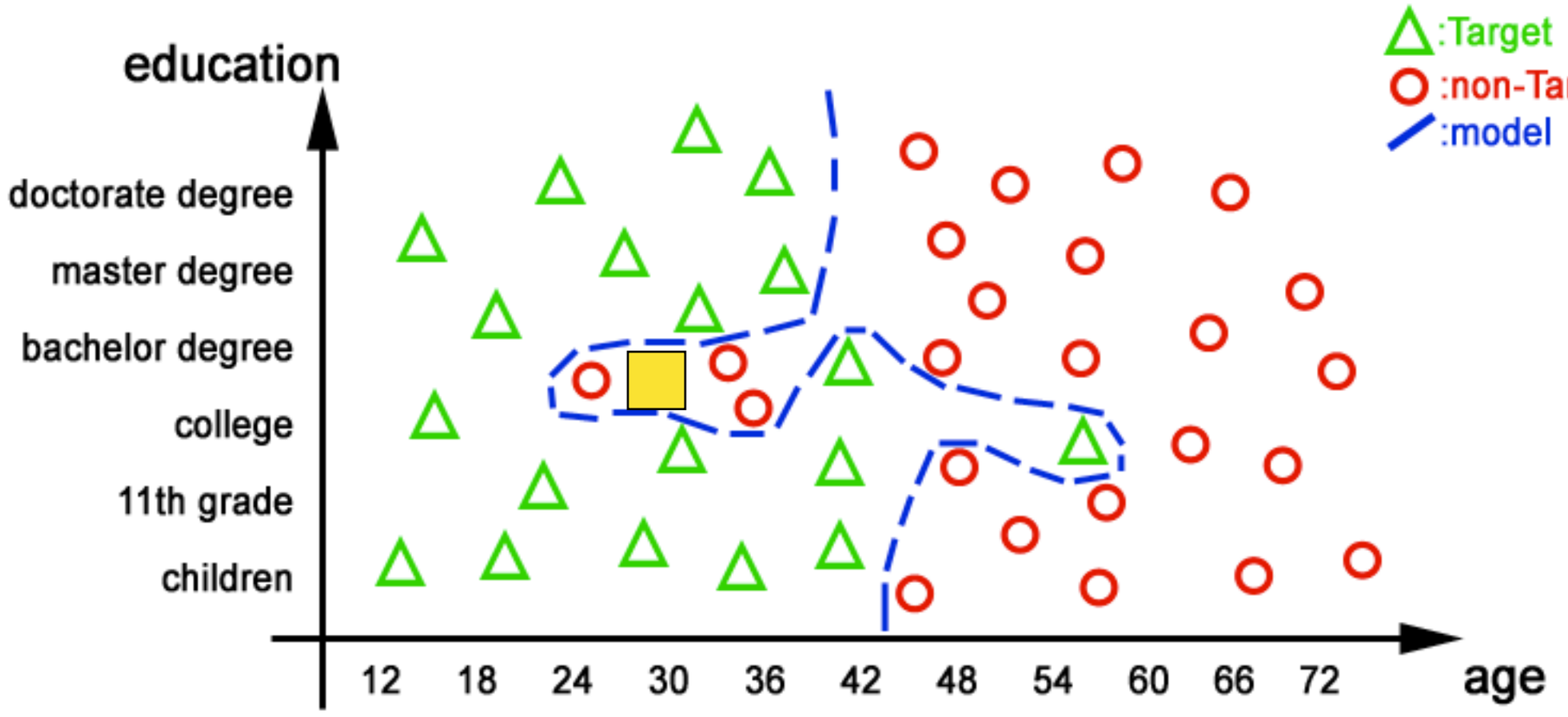
Model with errors



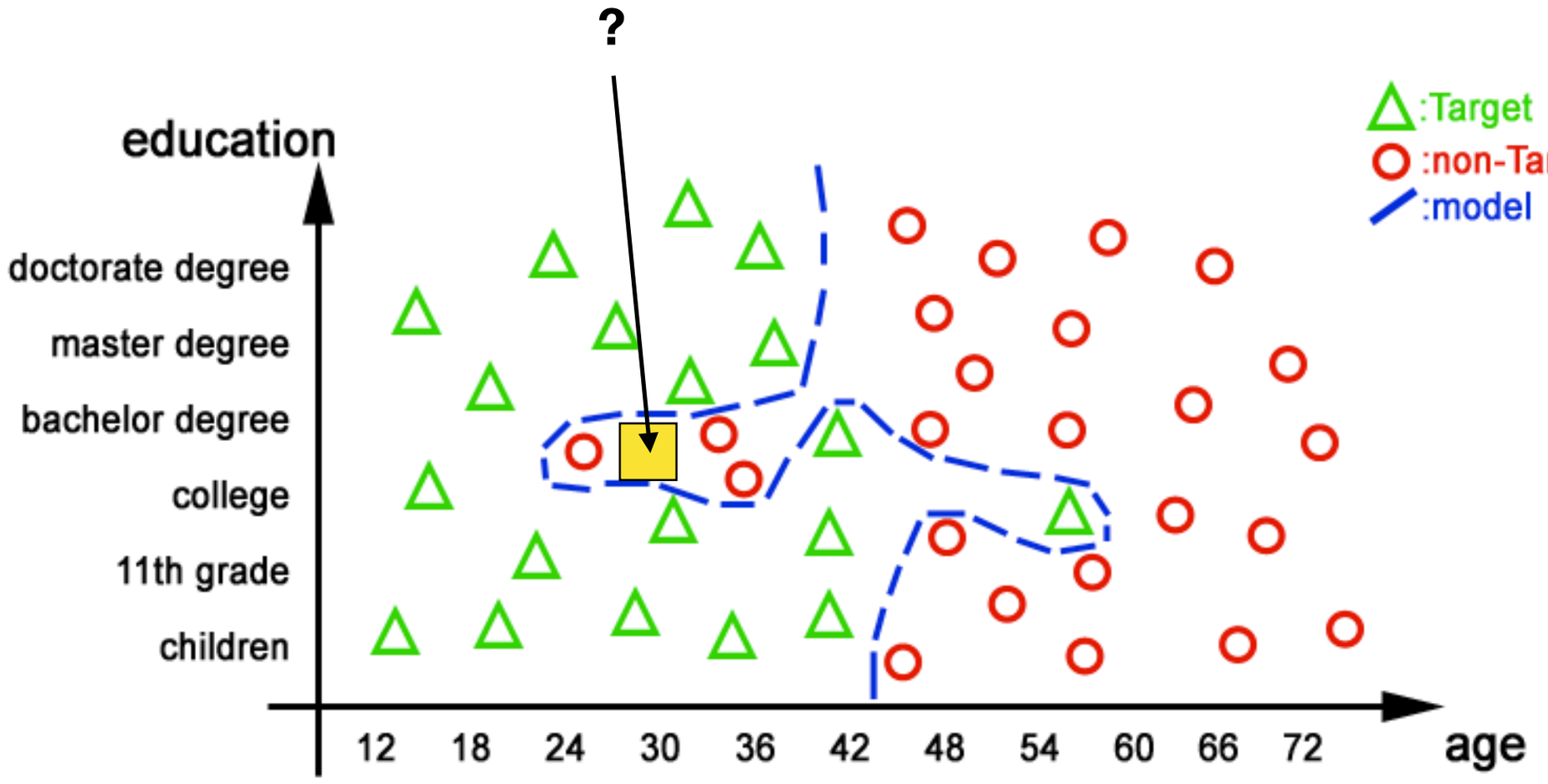
Overfitting



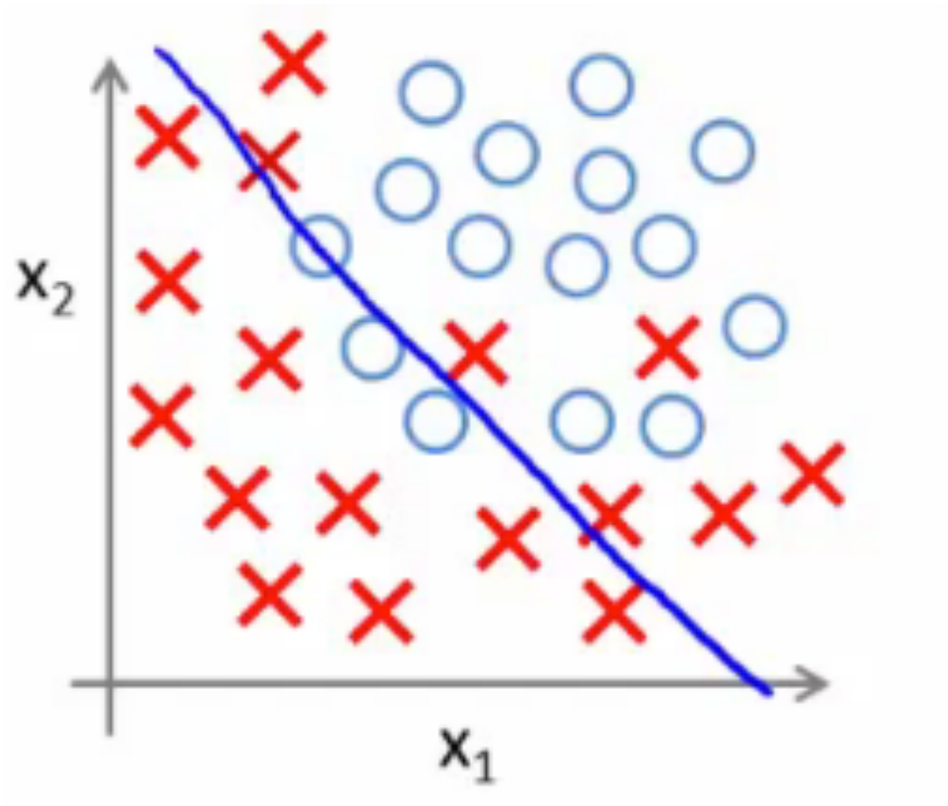
Overfitting



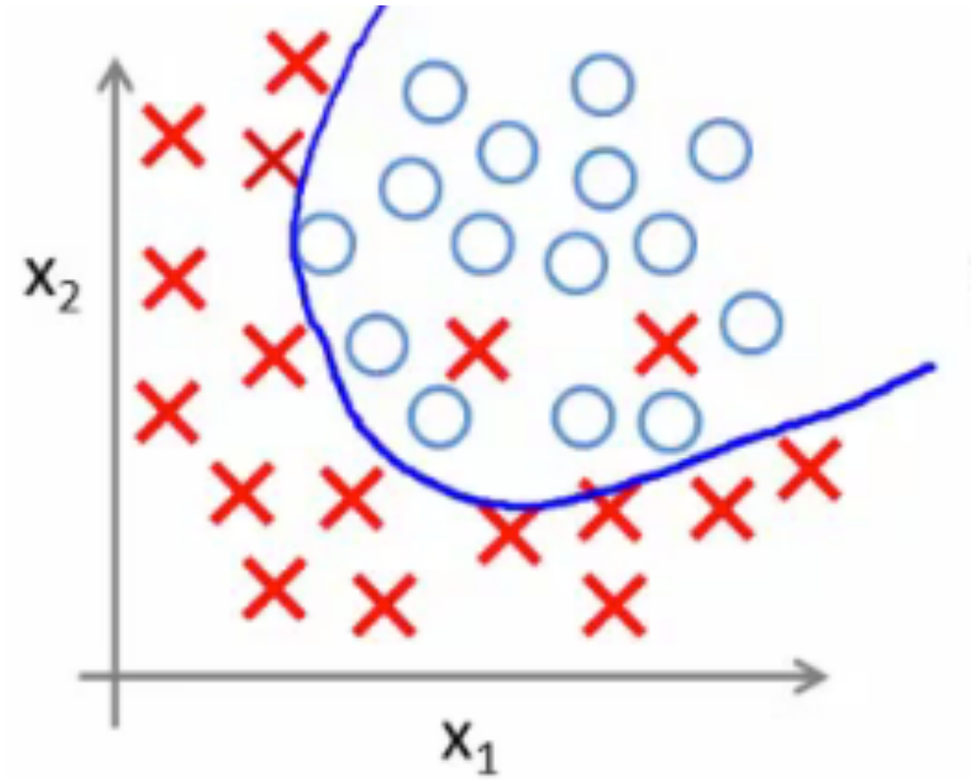
Overfitting



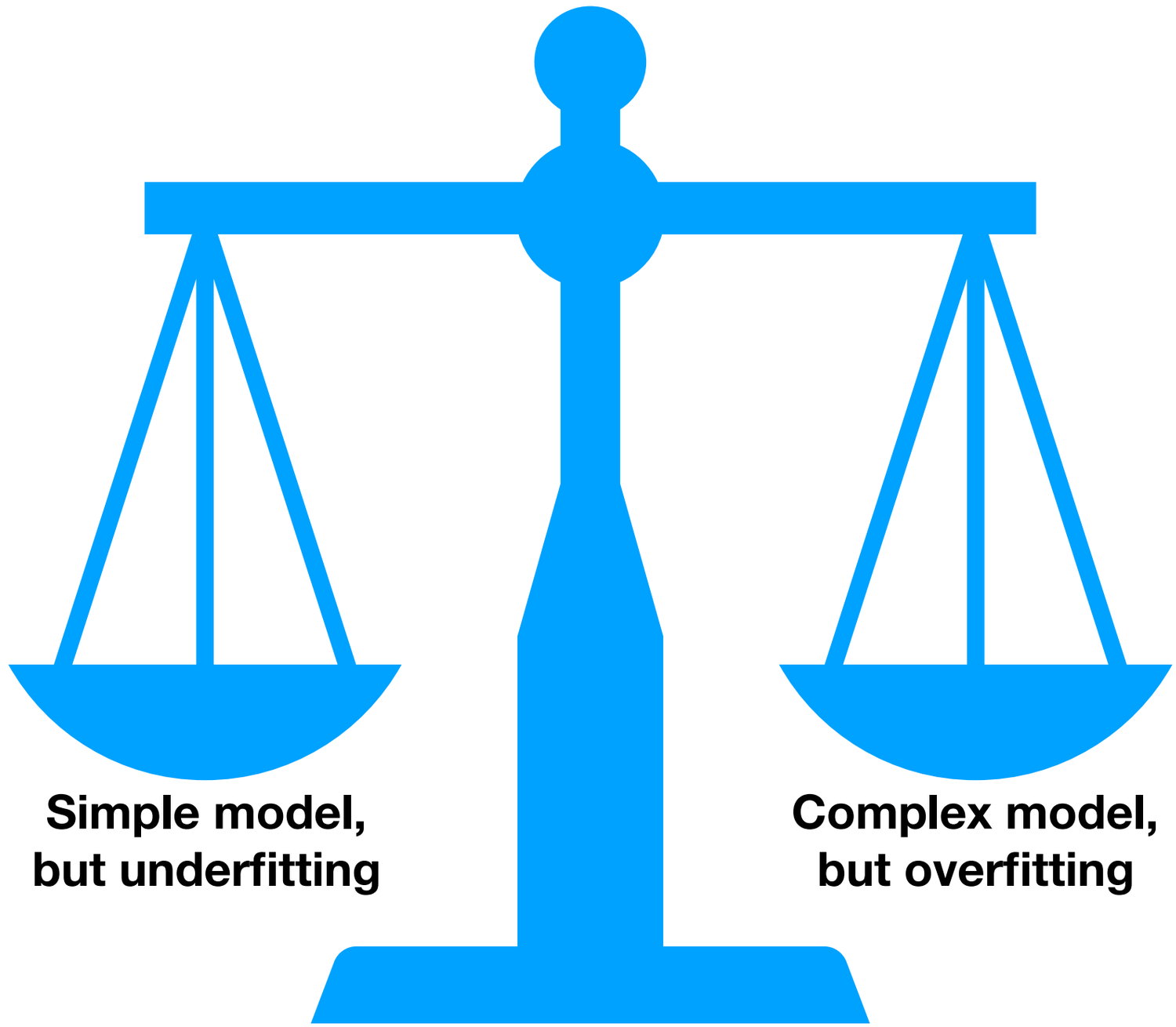
Underfitting



Underfitting

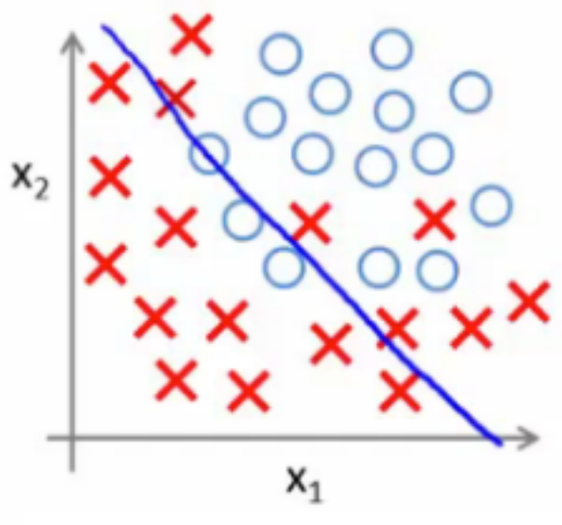


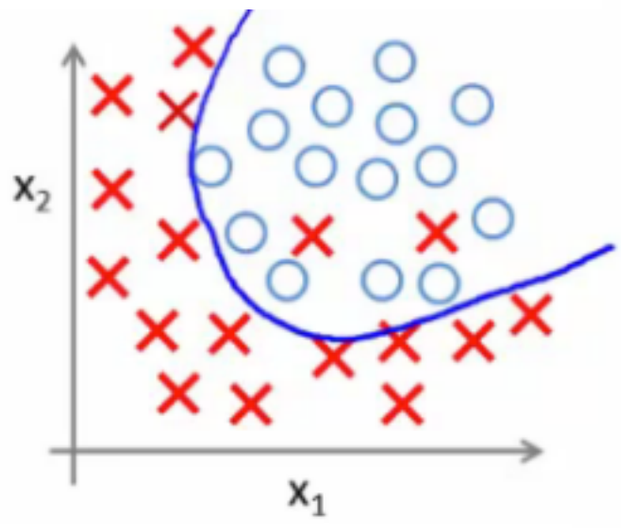
Best model

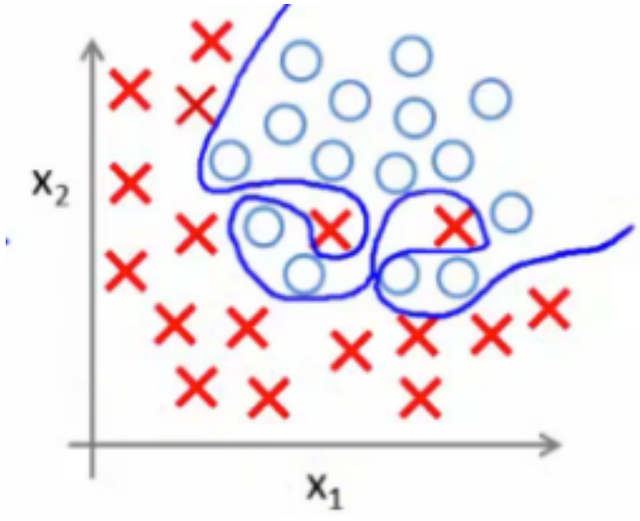

Simple model,
but underfitting
Complex model,
but overfitting

Best model

- How resolve Underfitting?
  - Check Errors
- How resolve Overfitting
  - ?







Train-Test Split

Train-Test Split

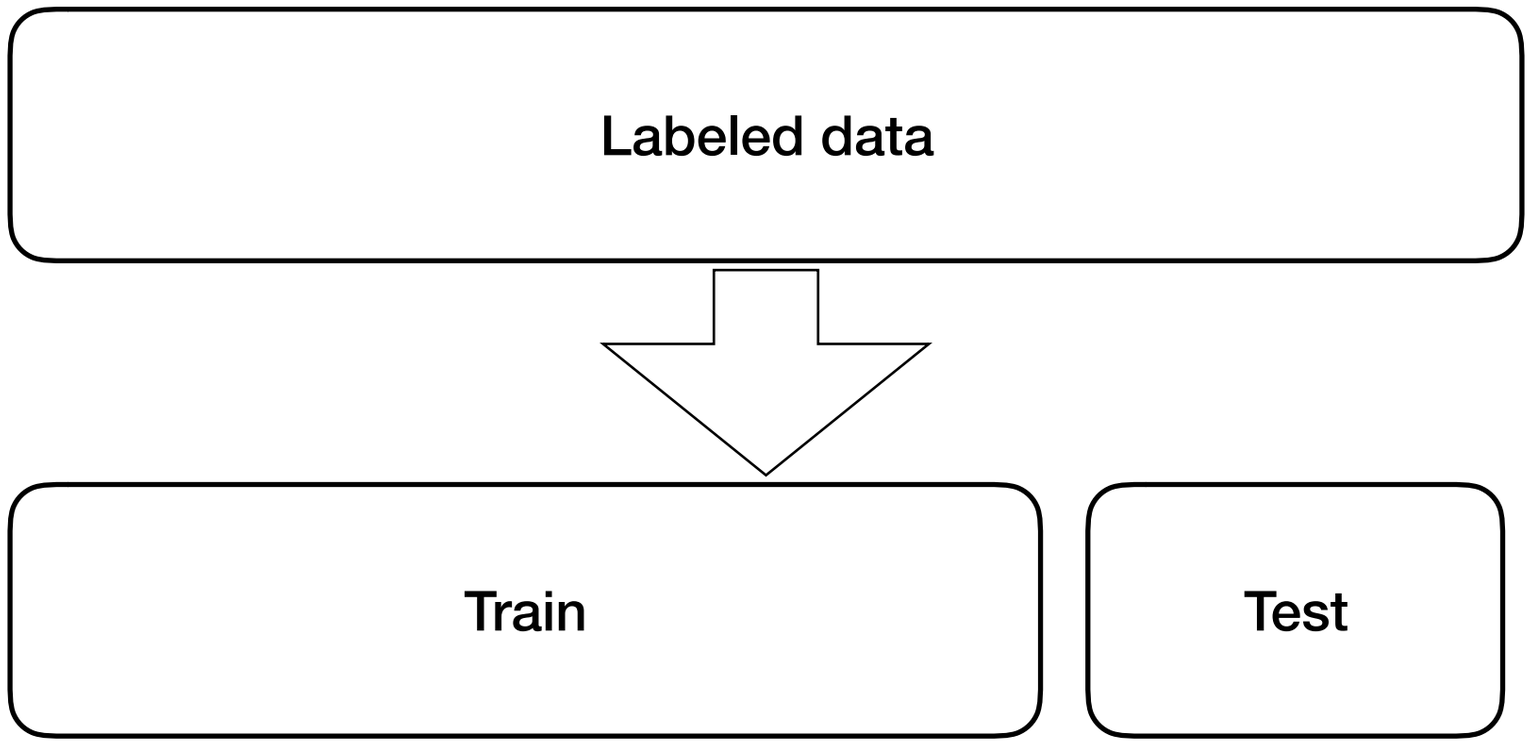
- Modelling of new elements

Train-Test Split

- Modelling of new elements

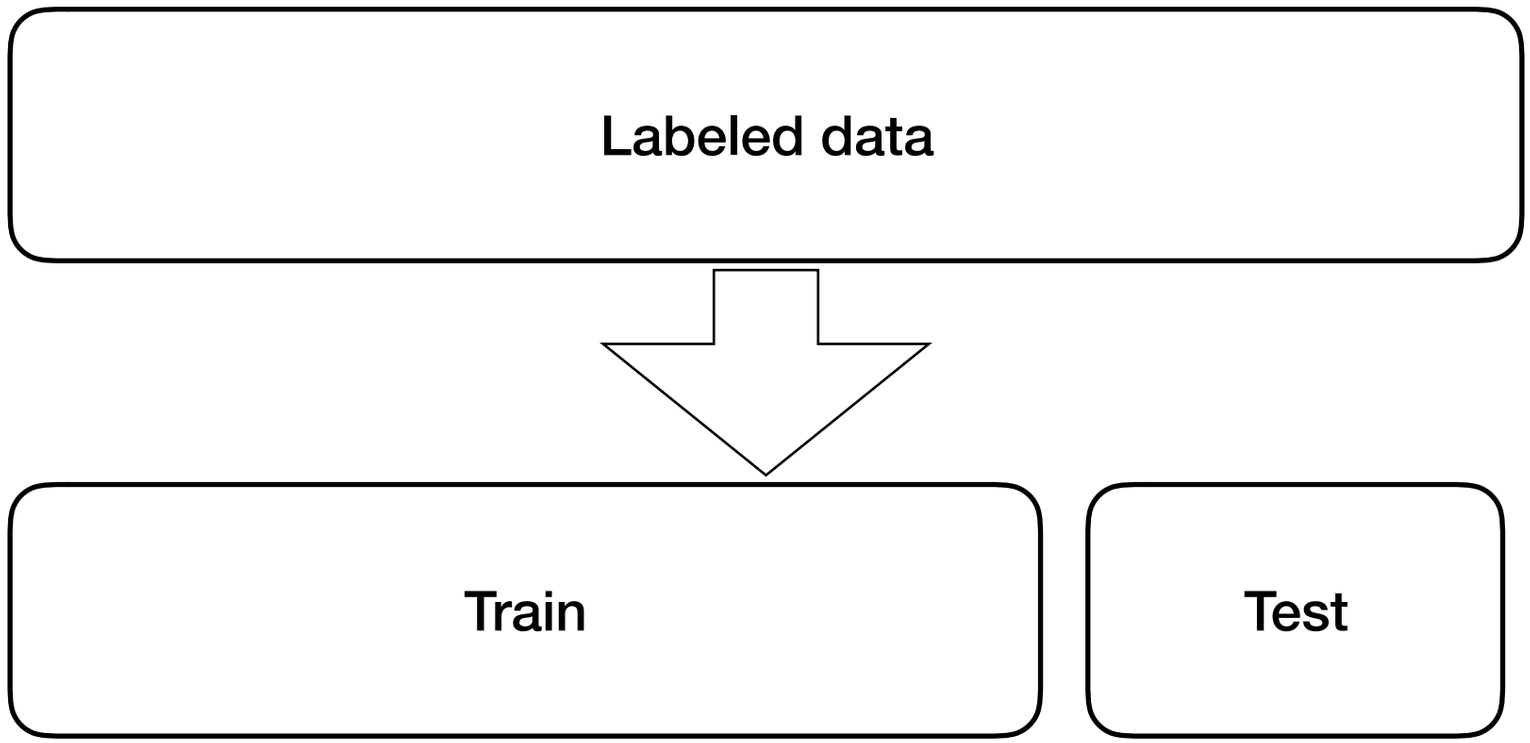
Train-Test Split

- Modelling of new elements

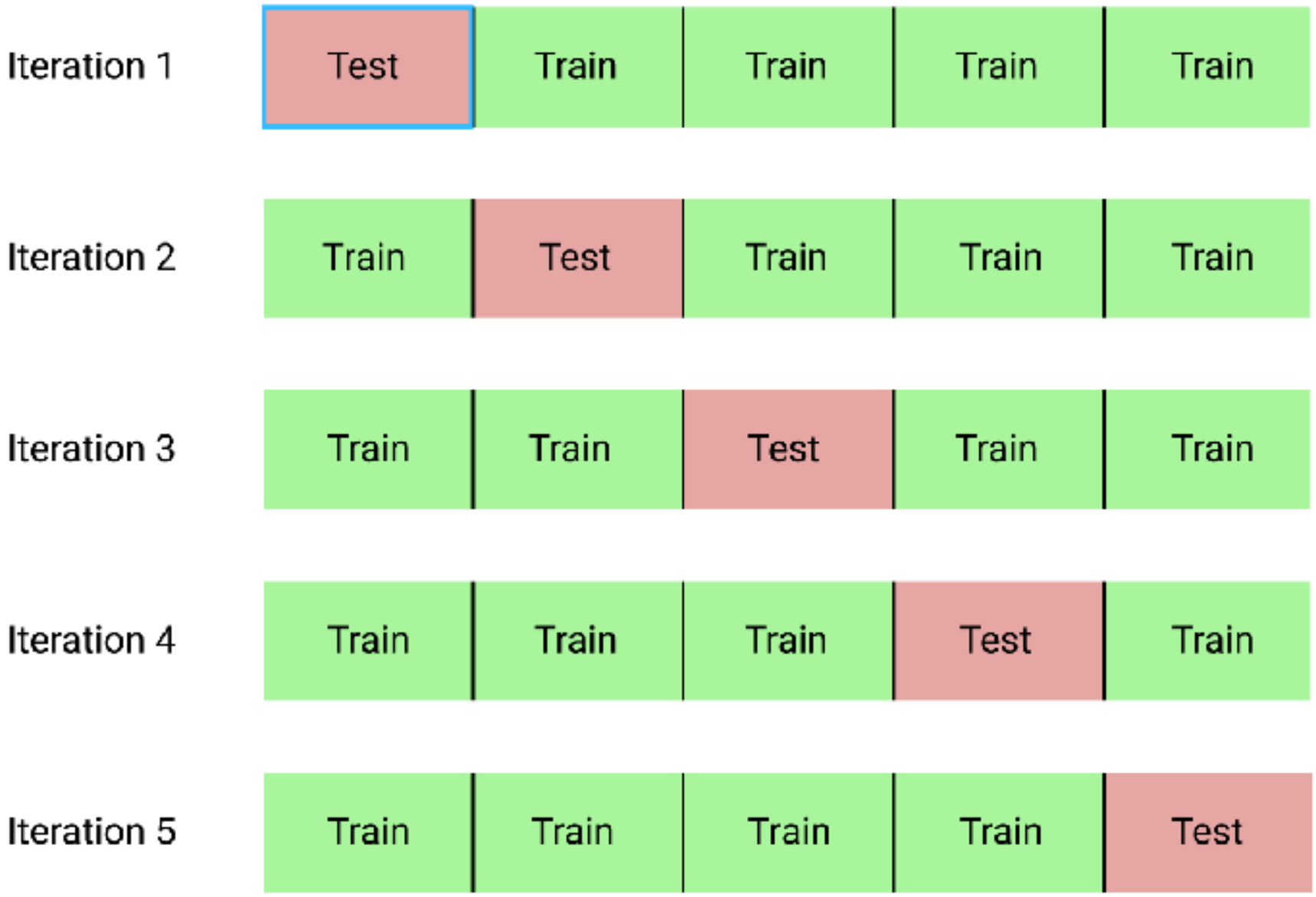


Holdout sample

- Modelling of new elements

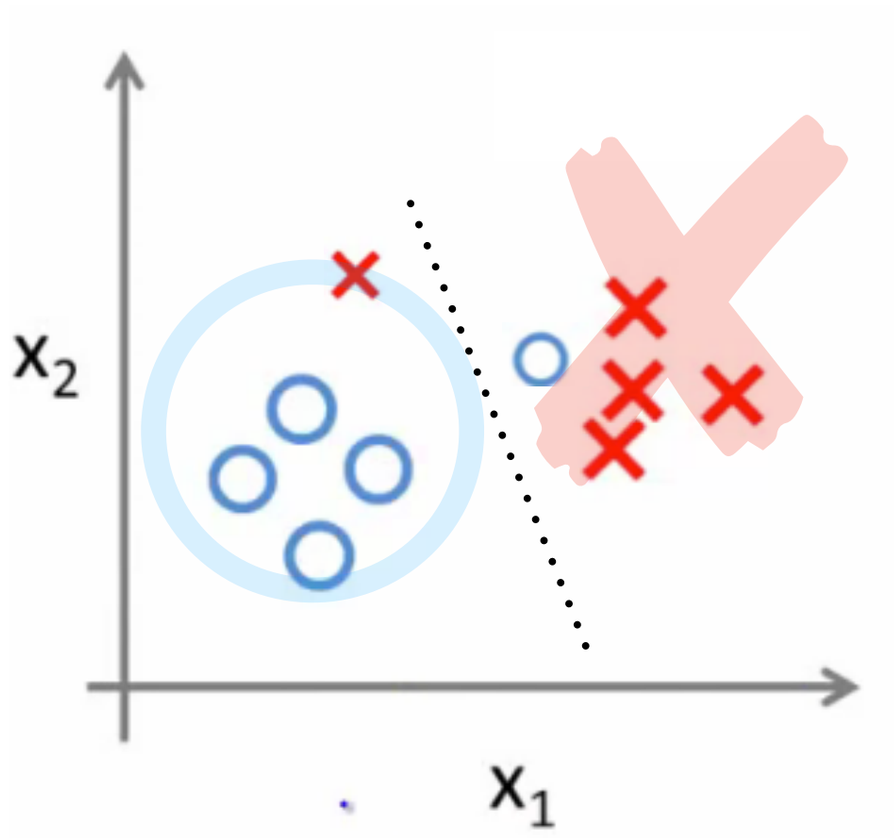


K-fold Cross Validation

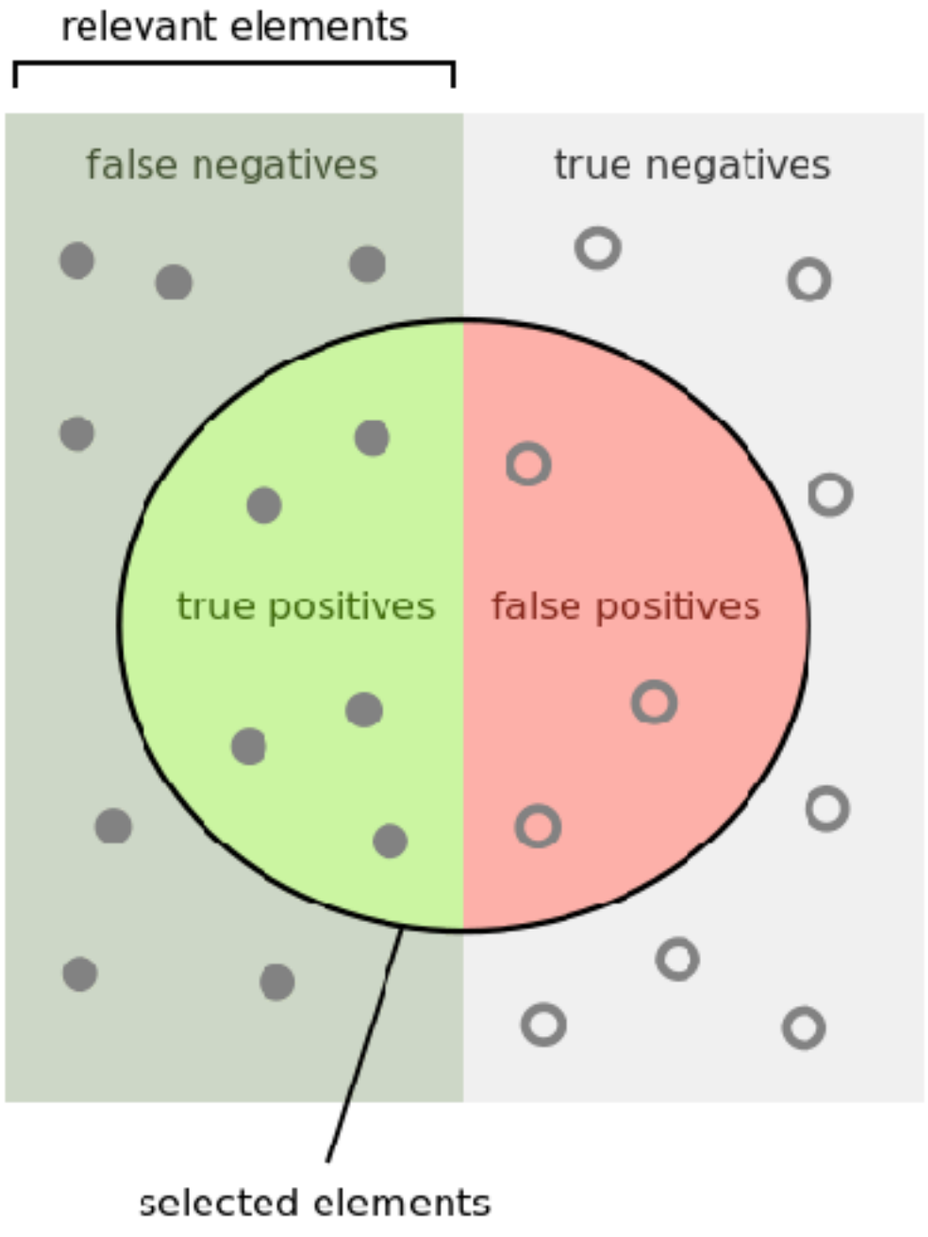


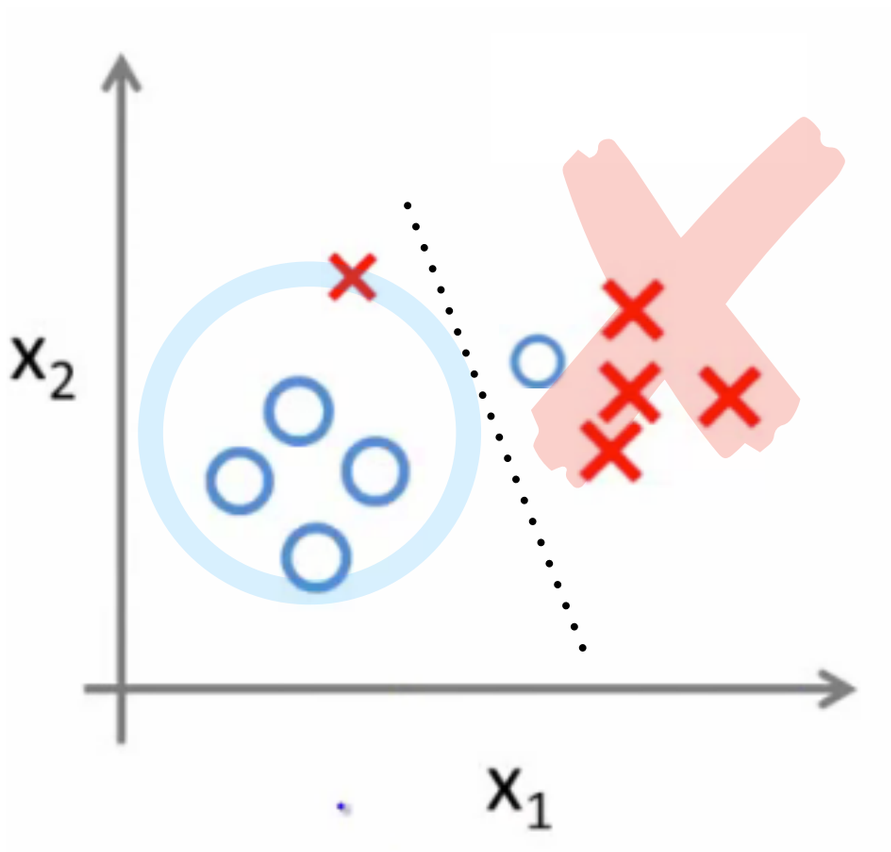
Classification metrics

		Prediction outcome: p	Prediction outcome: n
actual value	p′	True Positive	False Negative
	n′	False Positive	True Negative

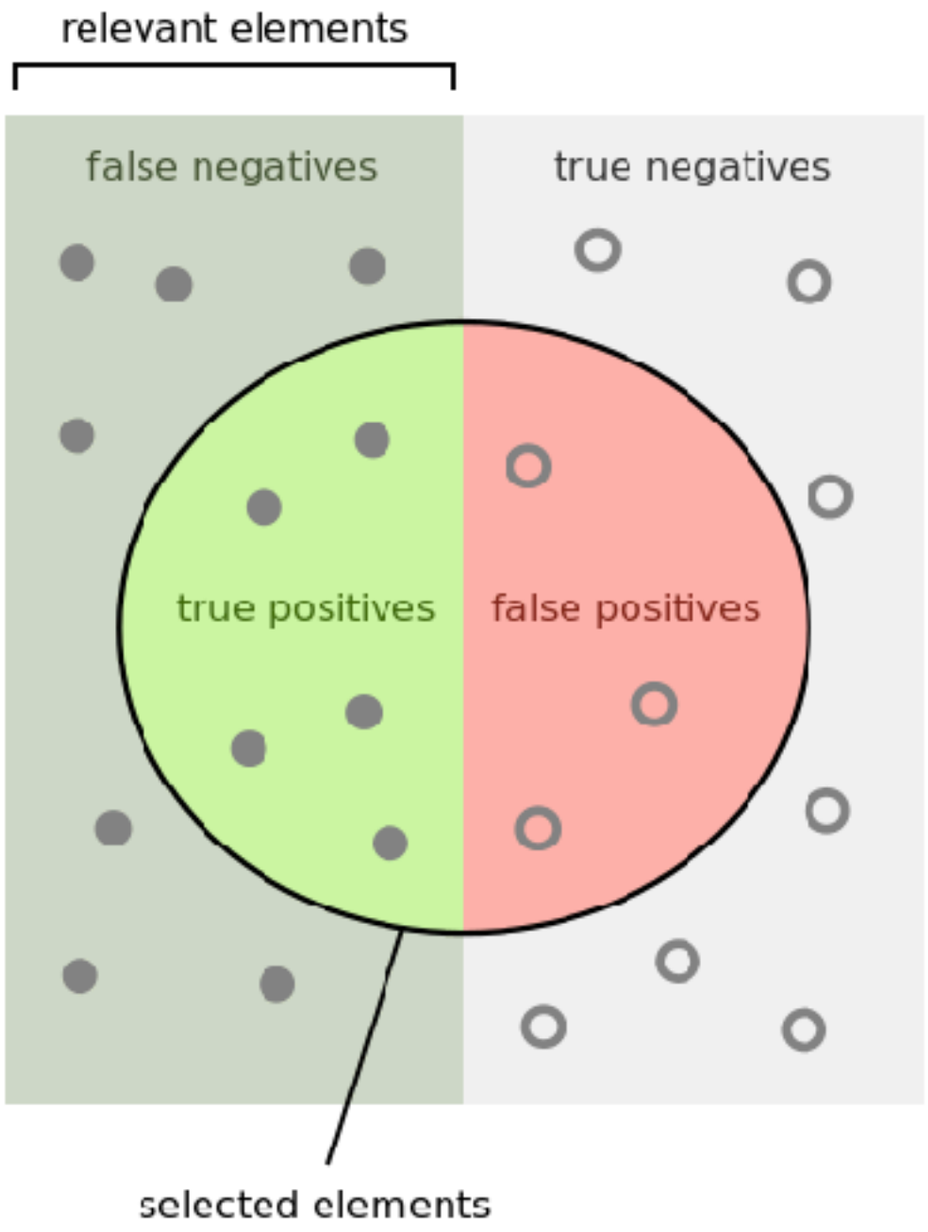


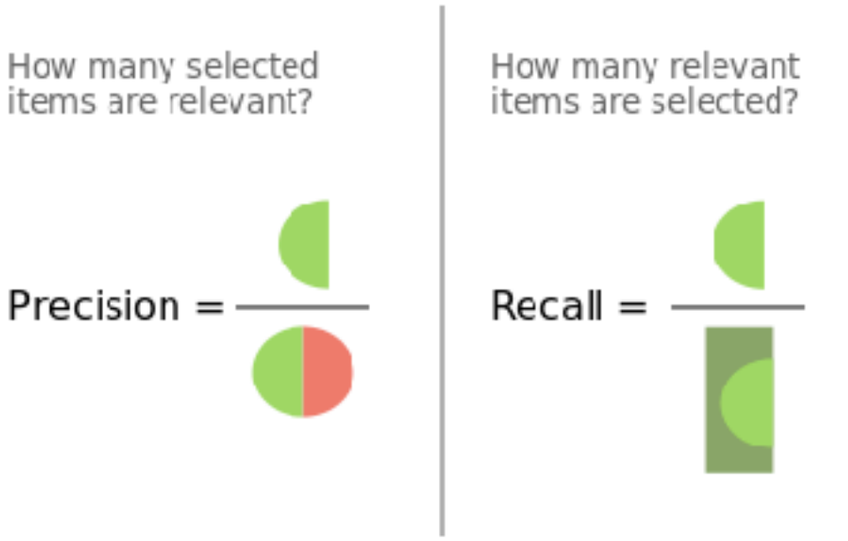
Classification metrics





Classification metrics





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

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

Classification metrics

- Combine **precision** and **recall**

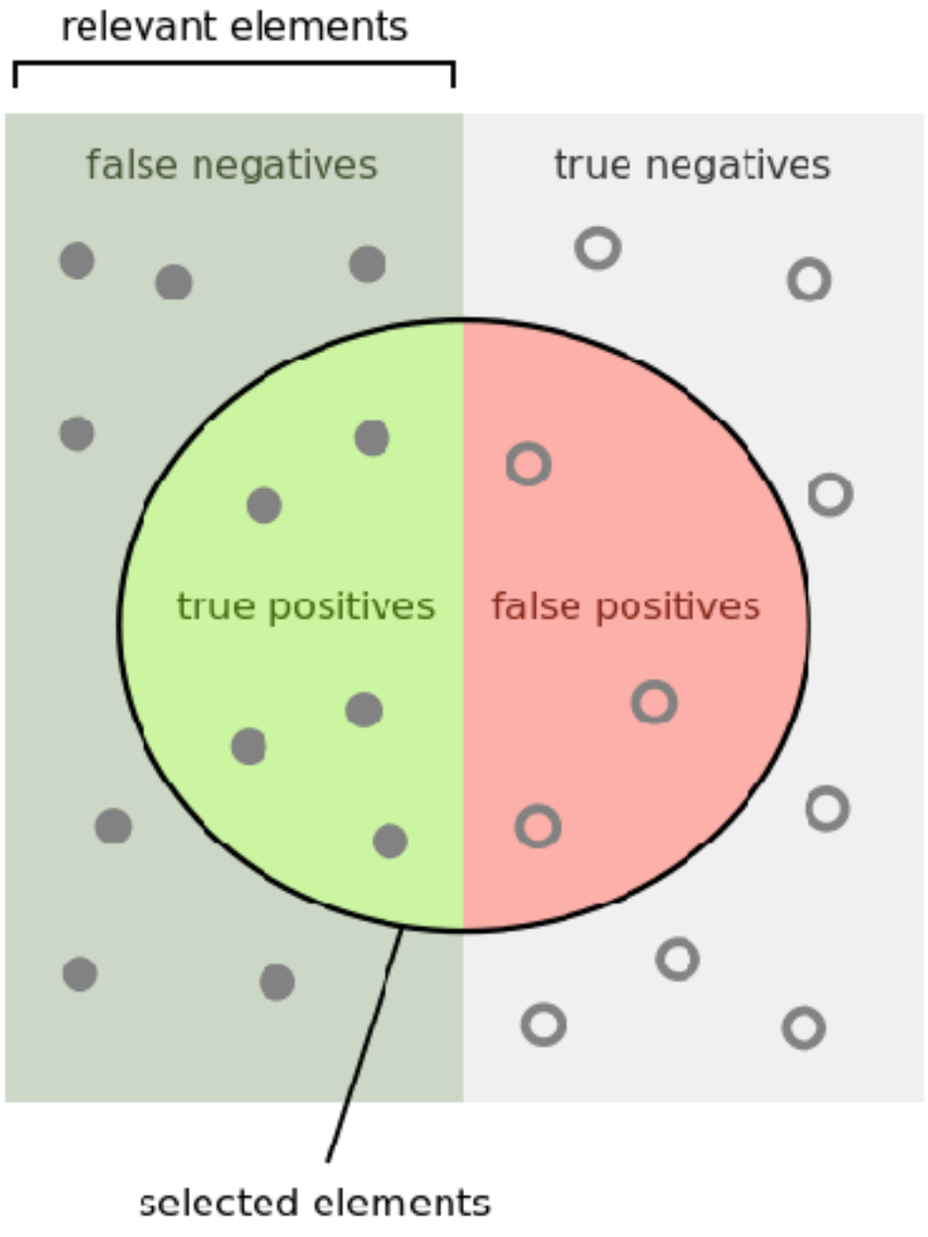
- F1-score
(harmonic average of the precision and recall)

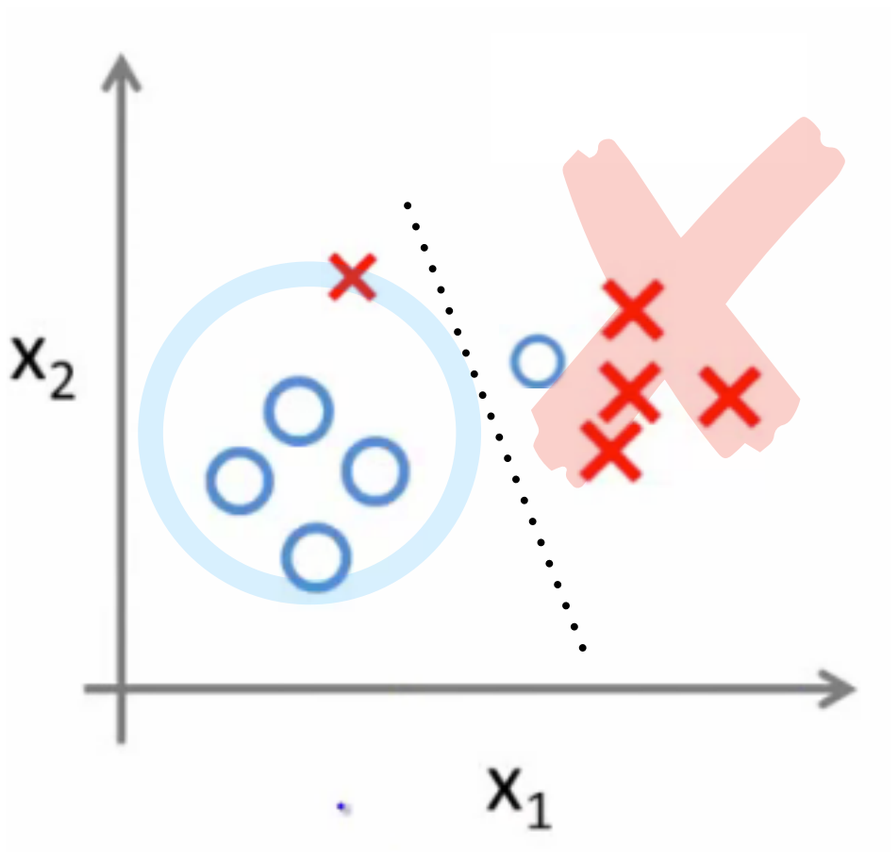
$$F_1 = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$$

- Fβ-score

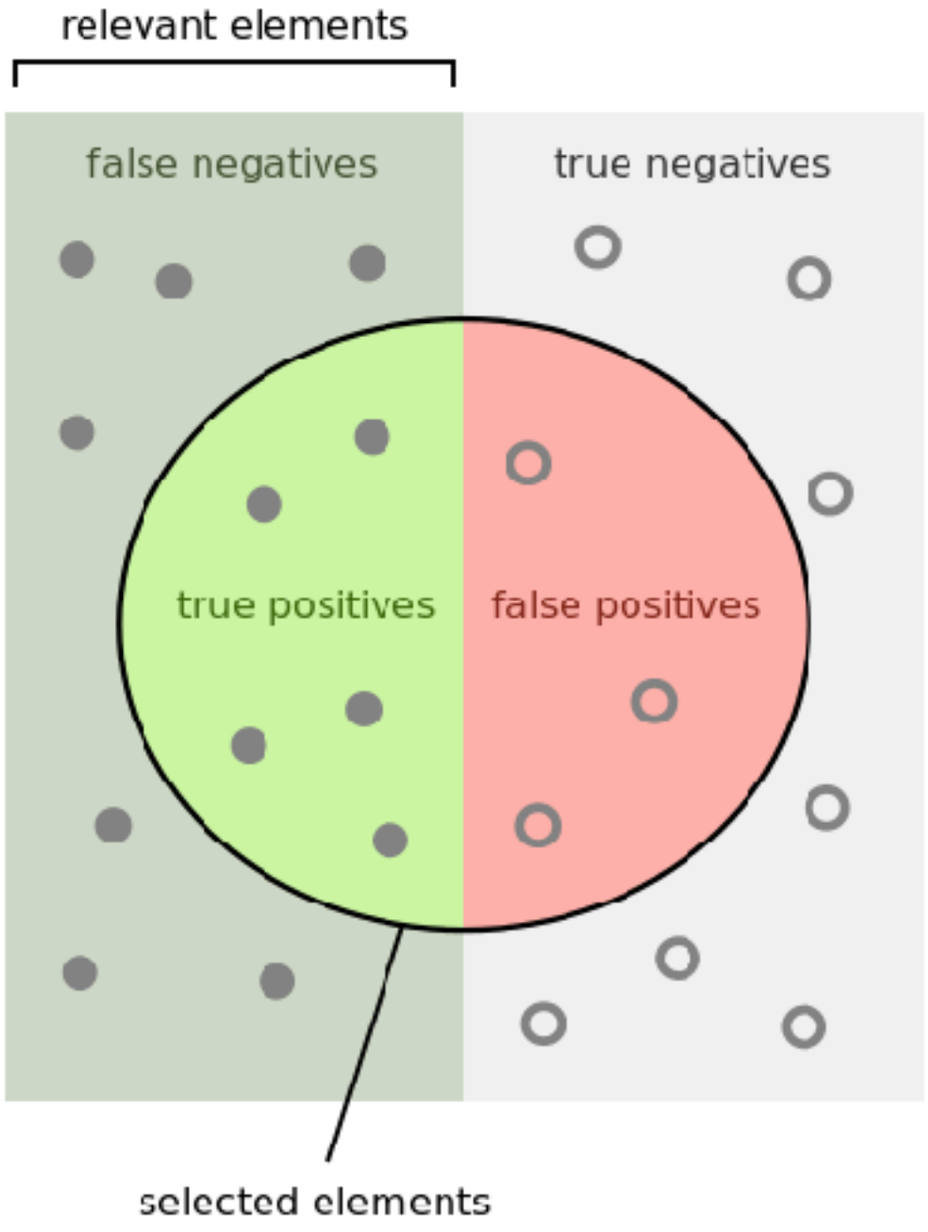
$$F_\beta = (1 + \beta^2) \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{(\beta^2 \cdot \mathrm{precision}) + \mathrm{recall}}.$$

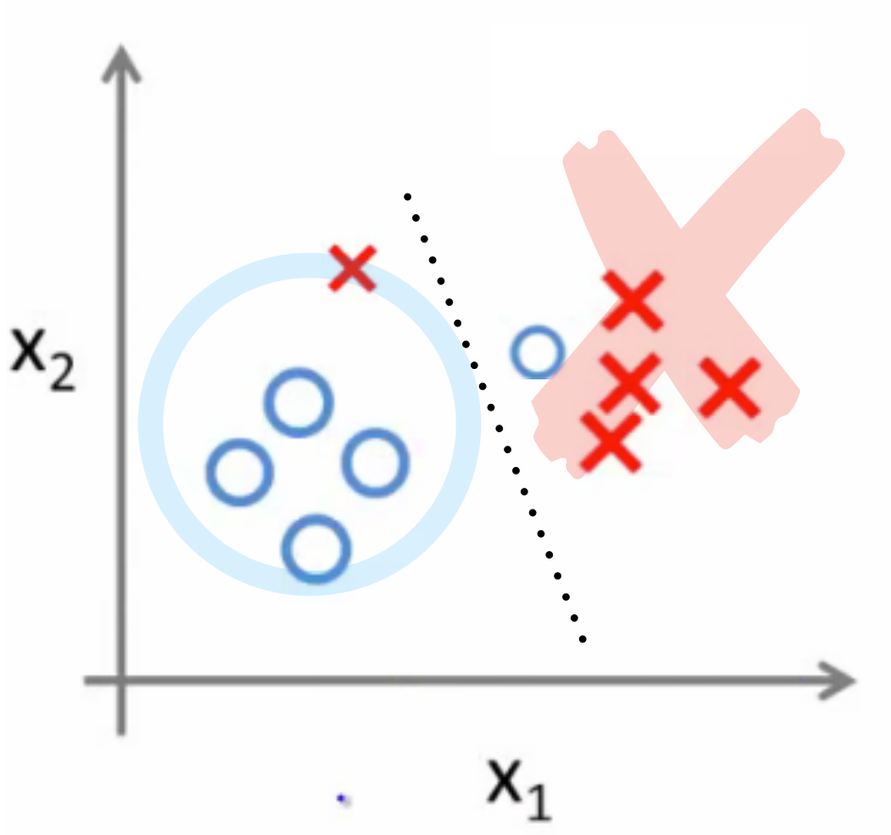
Classification metrics





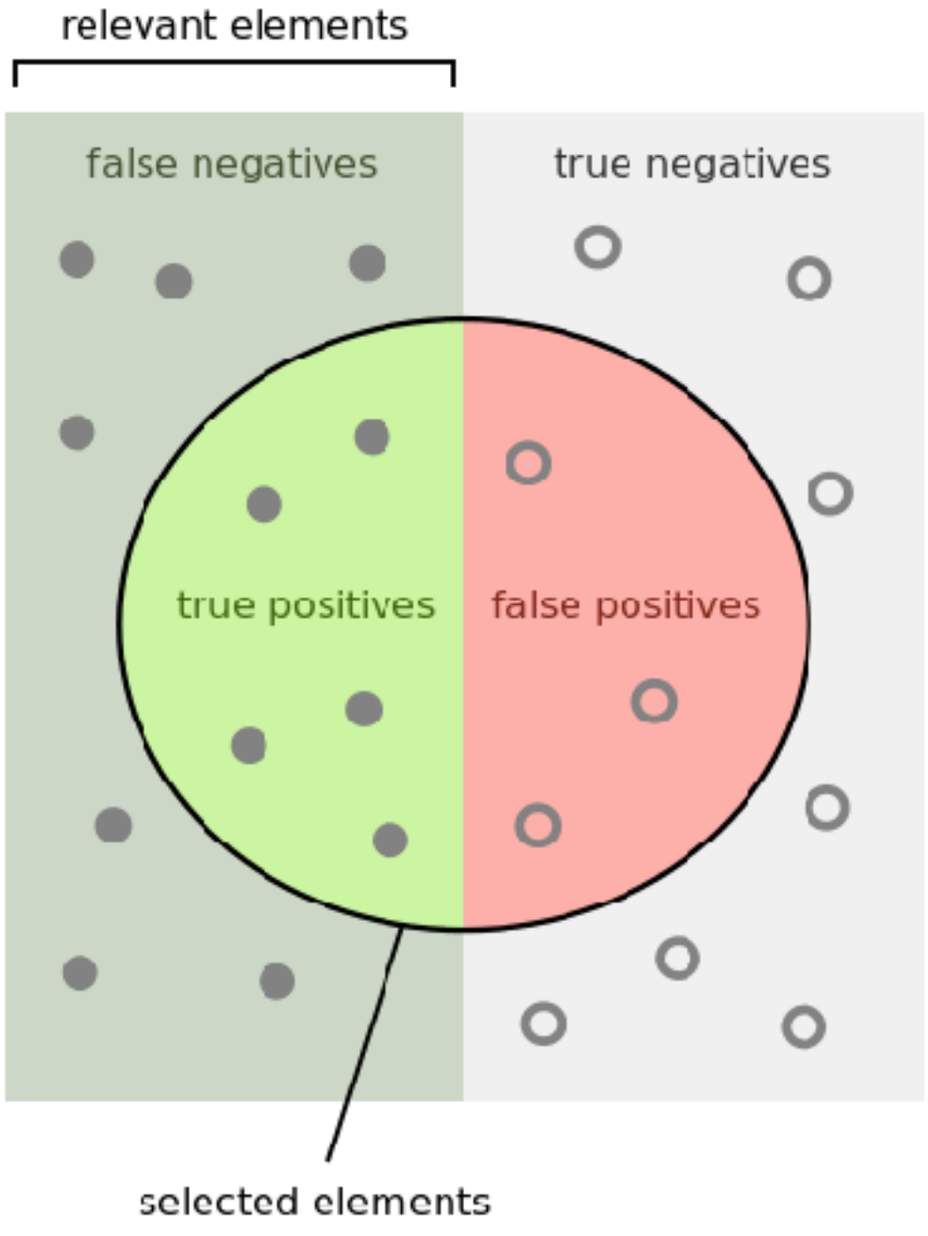
Classification metrics

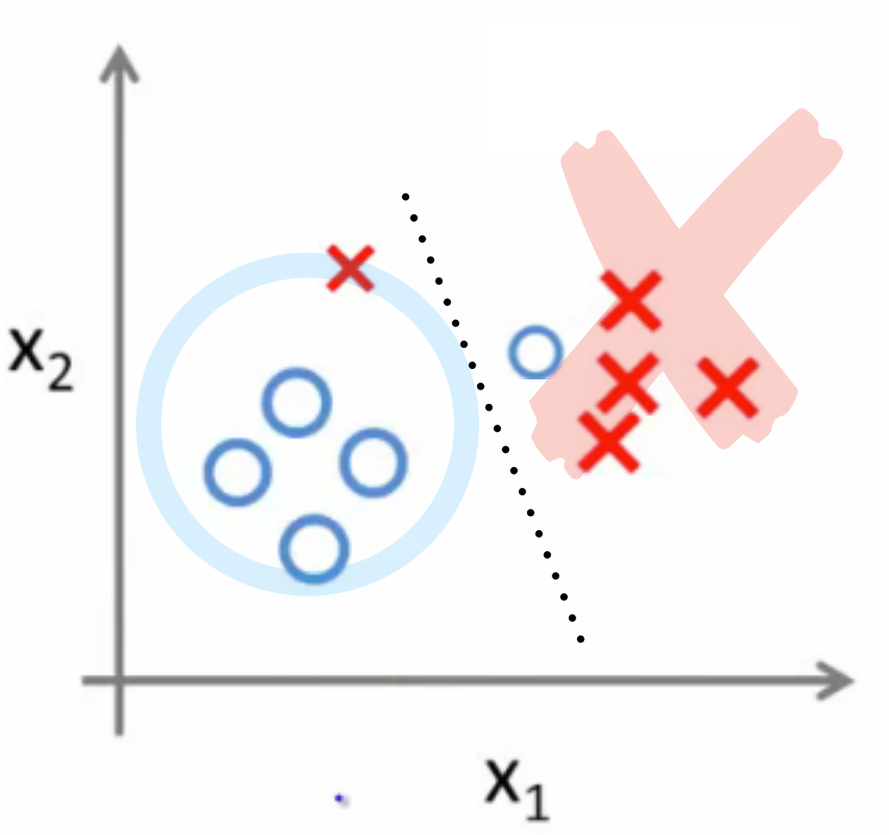




$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

Classification metrics





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

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

Confusion matrix

Prediction outcome		p	n
actual value	p′	True Positive	False Negative
	n′	False Positive	True Negative

		Predicted class	
		Cat	**Non-cat**
Actual class	**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

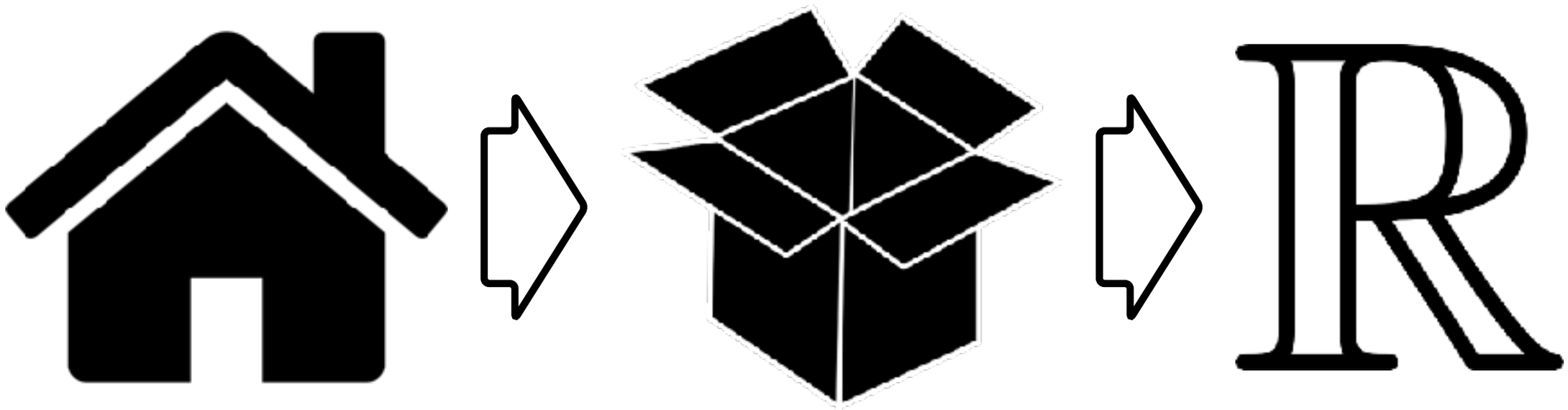
- True positive, False negative, True negative, False positive
- Precision, Recall
- F1-score/Fβ-score
- Accuracy
- True positive rate, False positive rate

Regression

Object

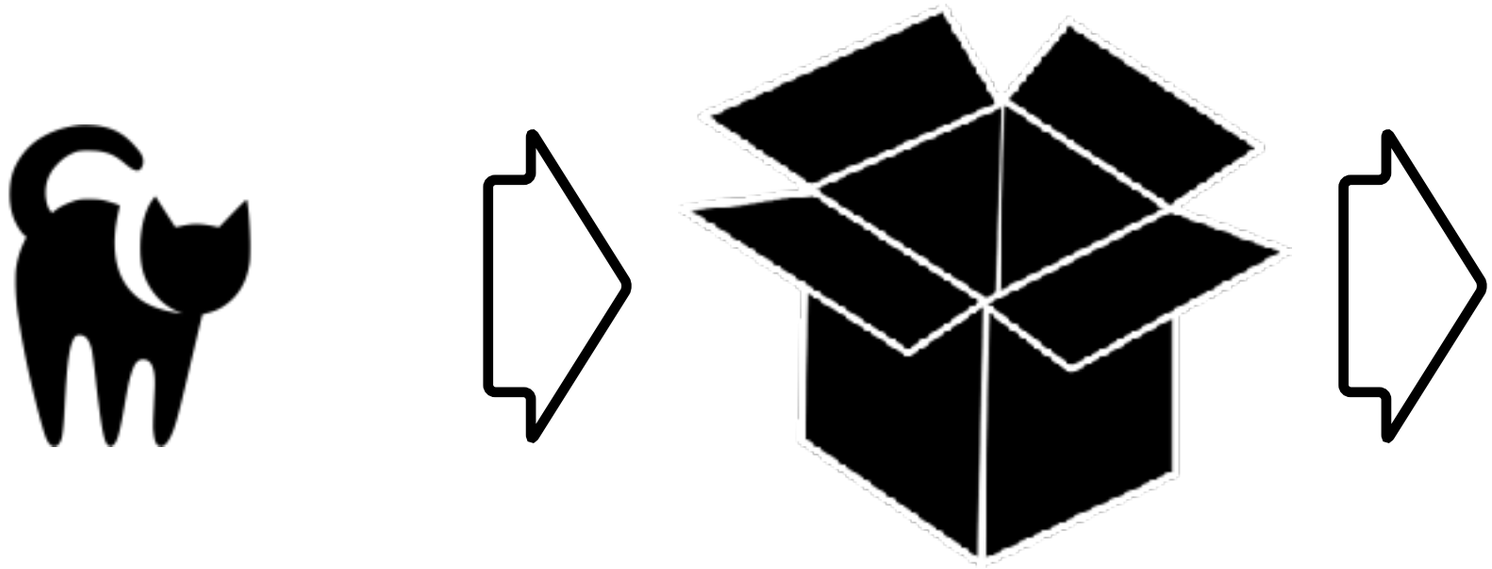
ML model

Real number



Binary-Classification

Object **ML model** **{0, 1}**



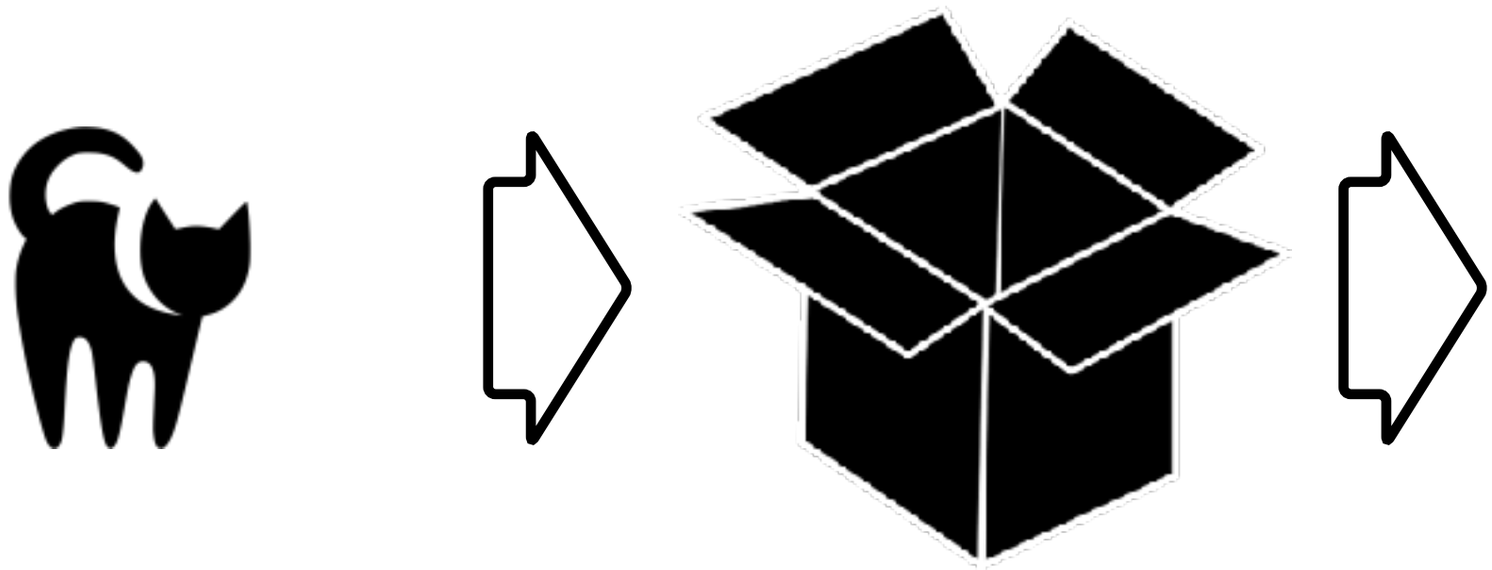
Cat
or
Not-cat

Binary-Classification

Object

ML model

Probability
between 0 and 1

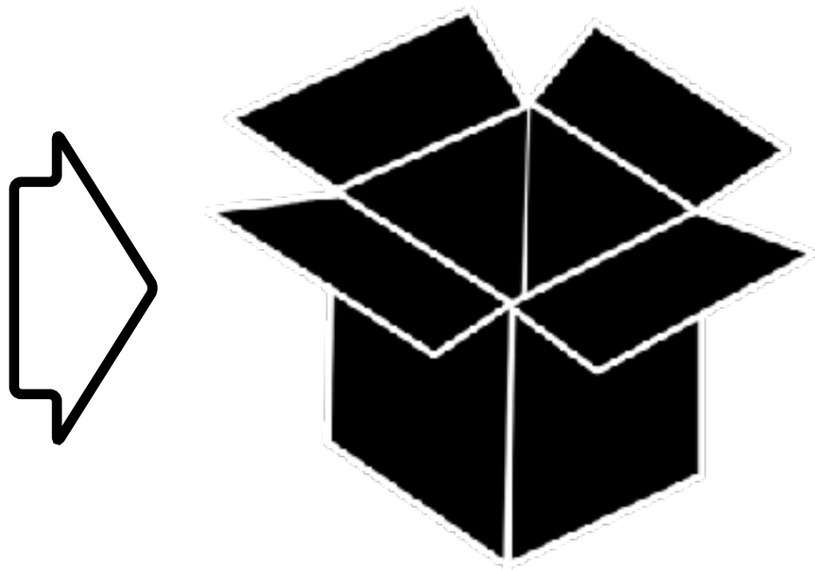


0.83

Binary-Classification

Object

ML model



Probability
between 0 and 1

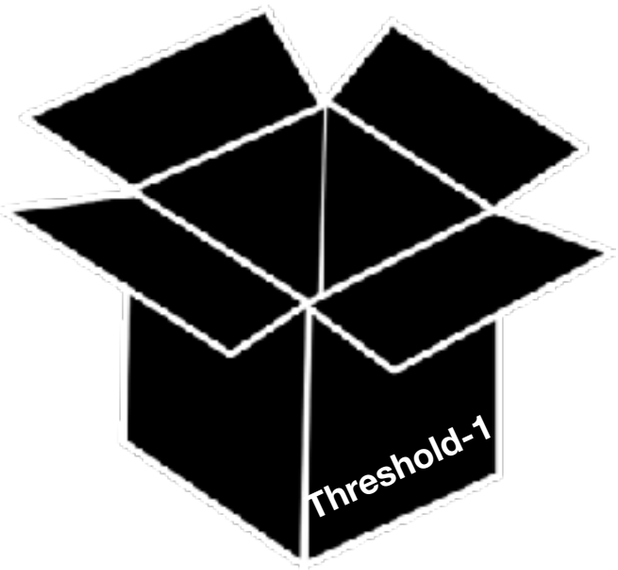
0.83

Threshold

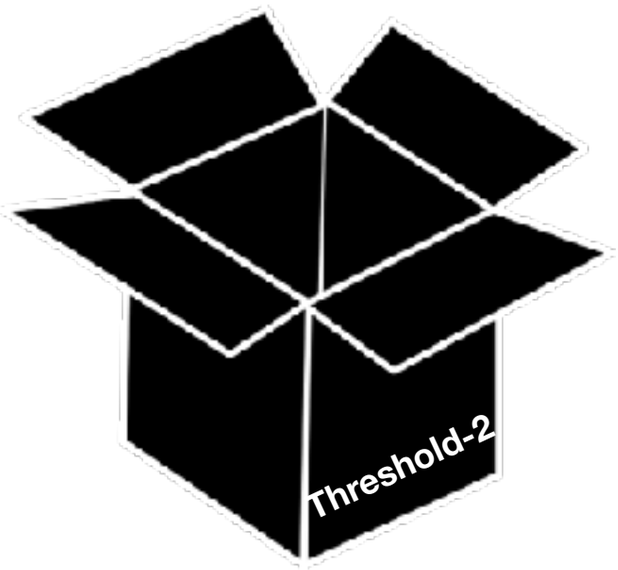
Cat
or
Not-cat

Compare two ML models

Model-1



Model-2



Precision-1

Recall-1

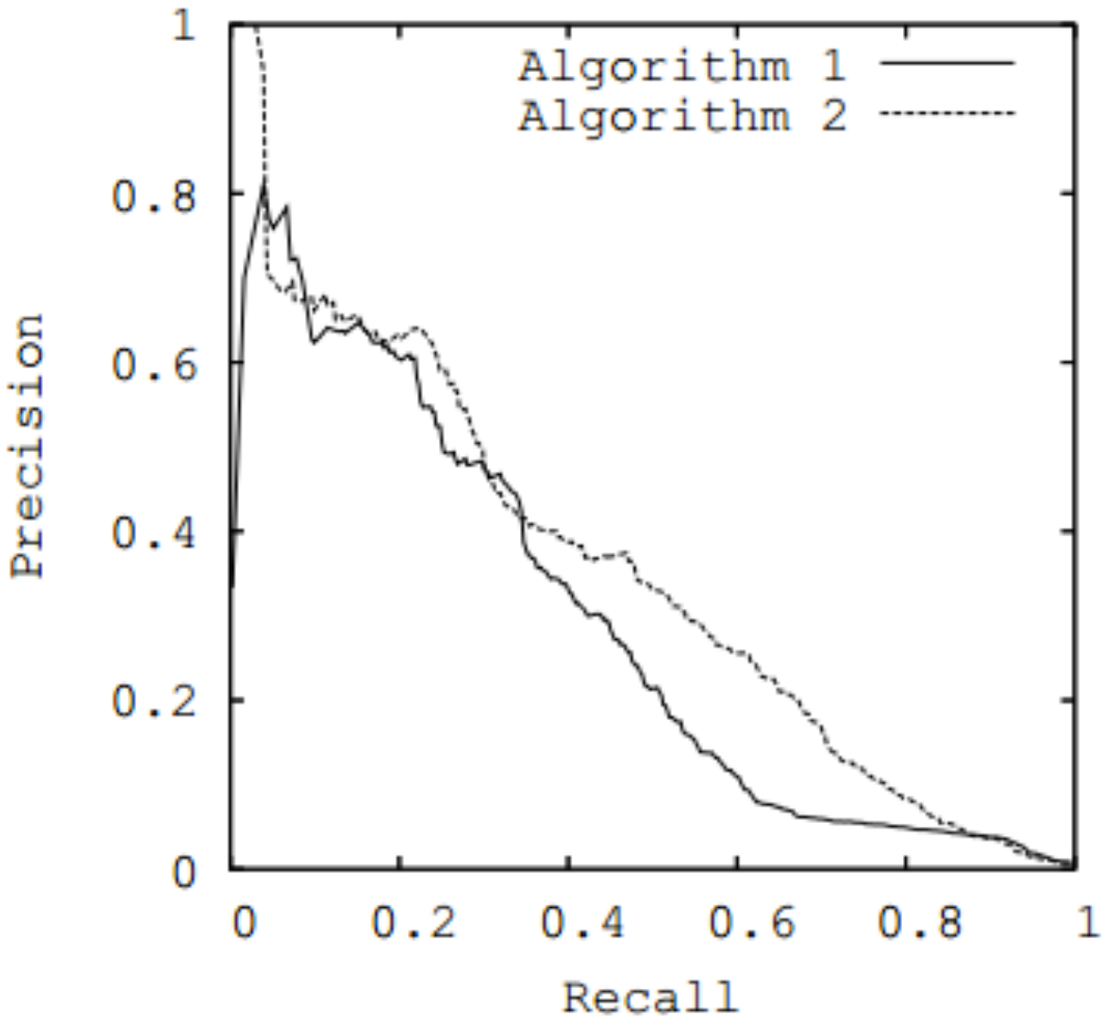
Precision-2

Recall-2

Threshold in classification

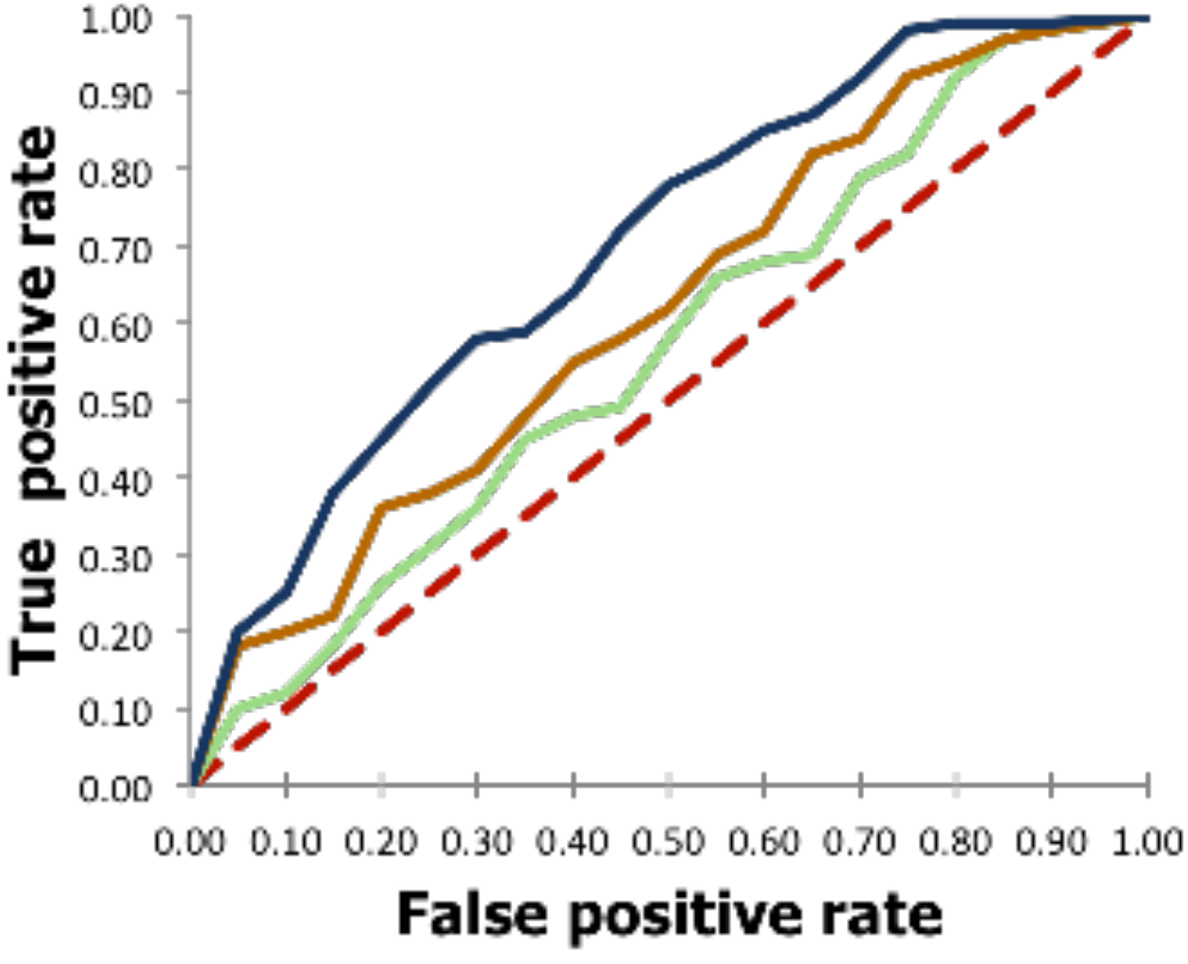
- How to choose the best model without binding to the threshold?
- Choose the same threshold?
  NO
- Need compare models without fixing the threshold!

Without fixing threshold



Precision-Recall curve

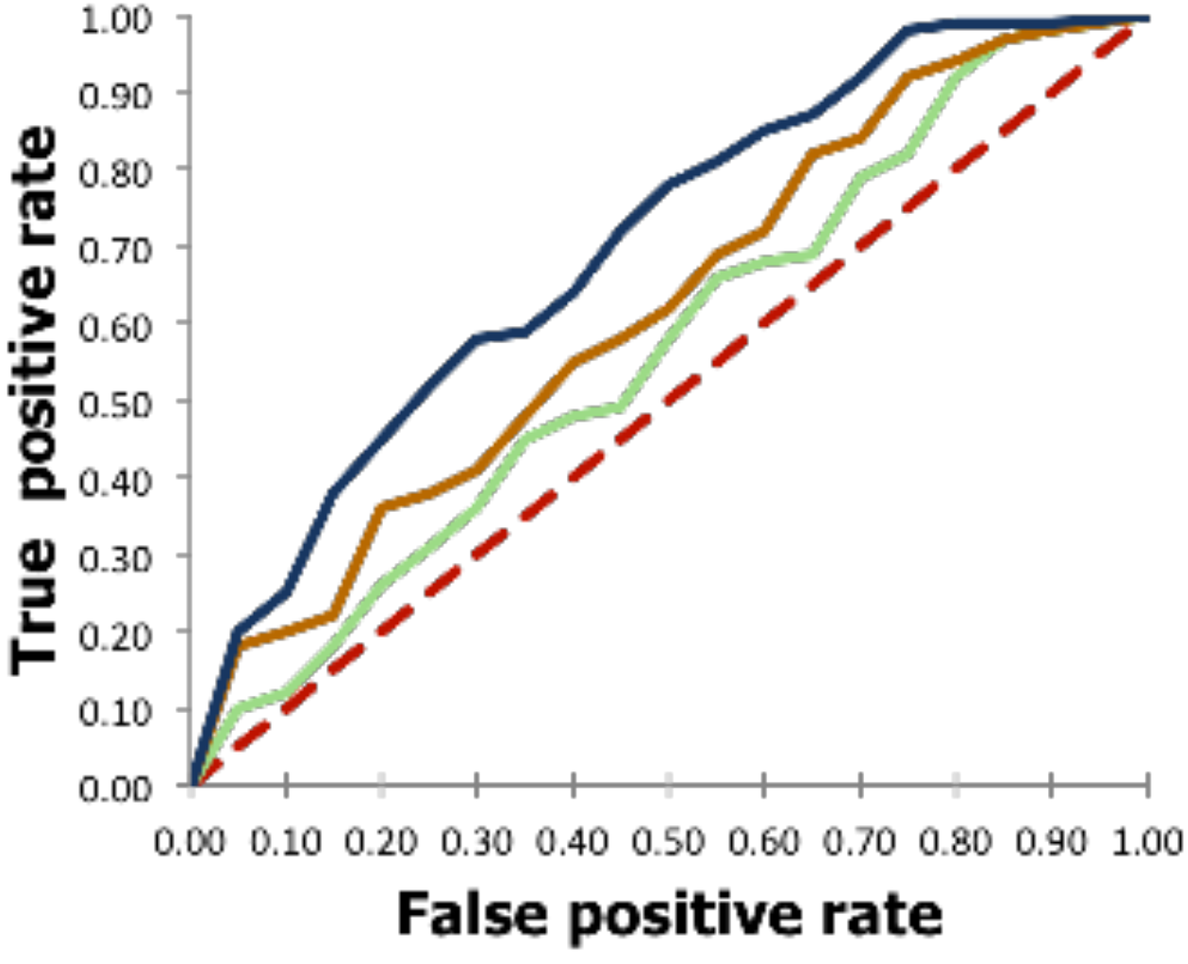
Without fixing threshold



ROC curve

(receiver operating characteristic)

Without fixing threshold



ROC curve

(receiver operating characteristic)

AUC ROC

(area under curve)

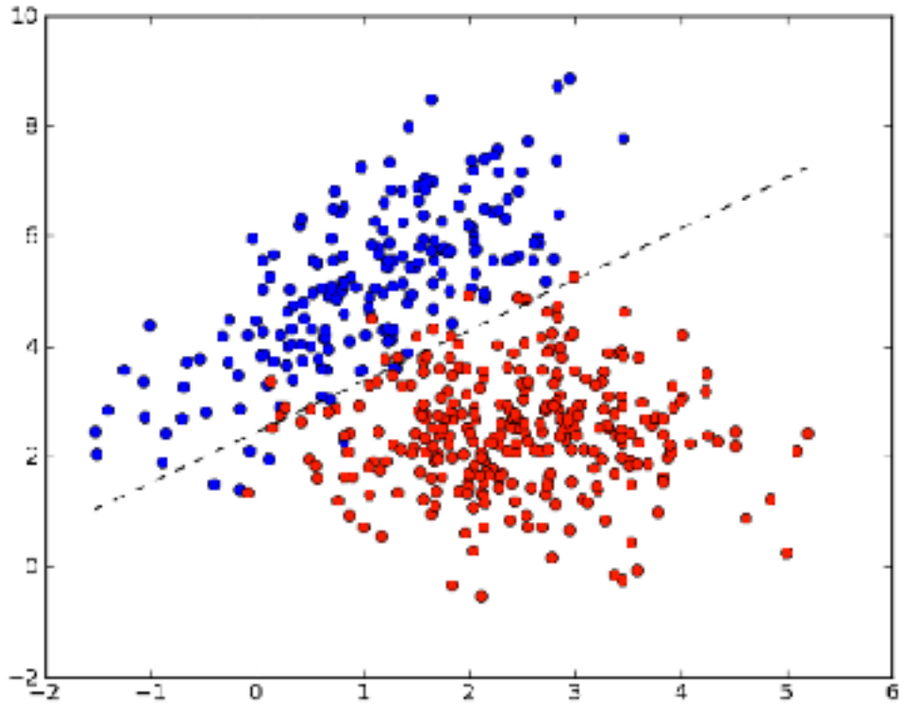
Without fixing threshold

LogLoss

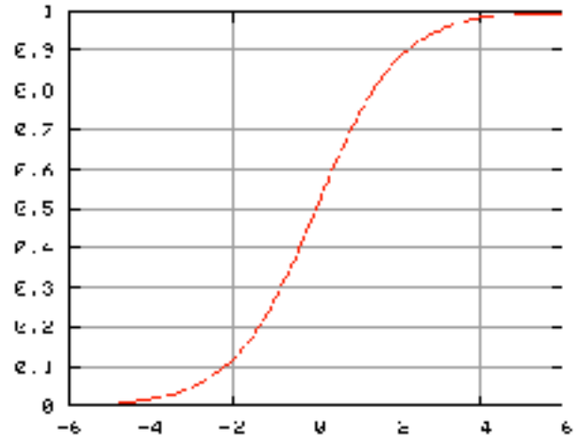
$$\text{LogLoss} = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right]$$

Logistic regression

- *f(X) = a*X1 + b*X2 + c*X3 + d - linear regression example*



- Sigmoid $f(z) = \frac{1}{1 + e^{-z}}$

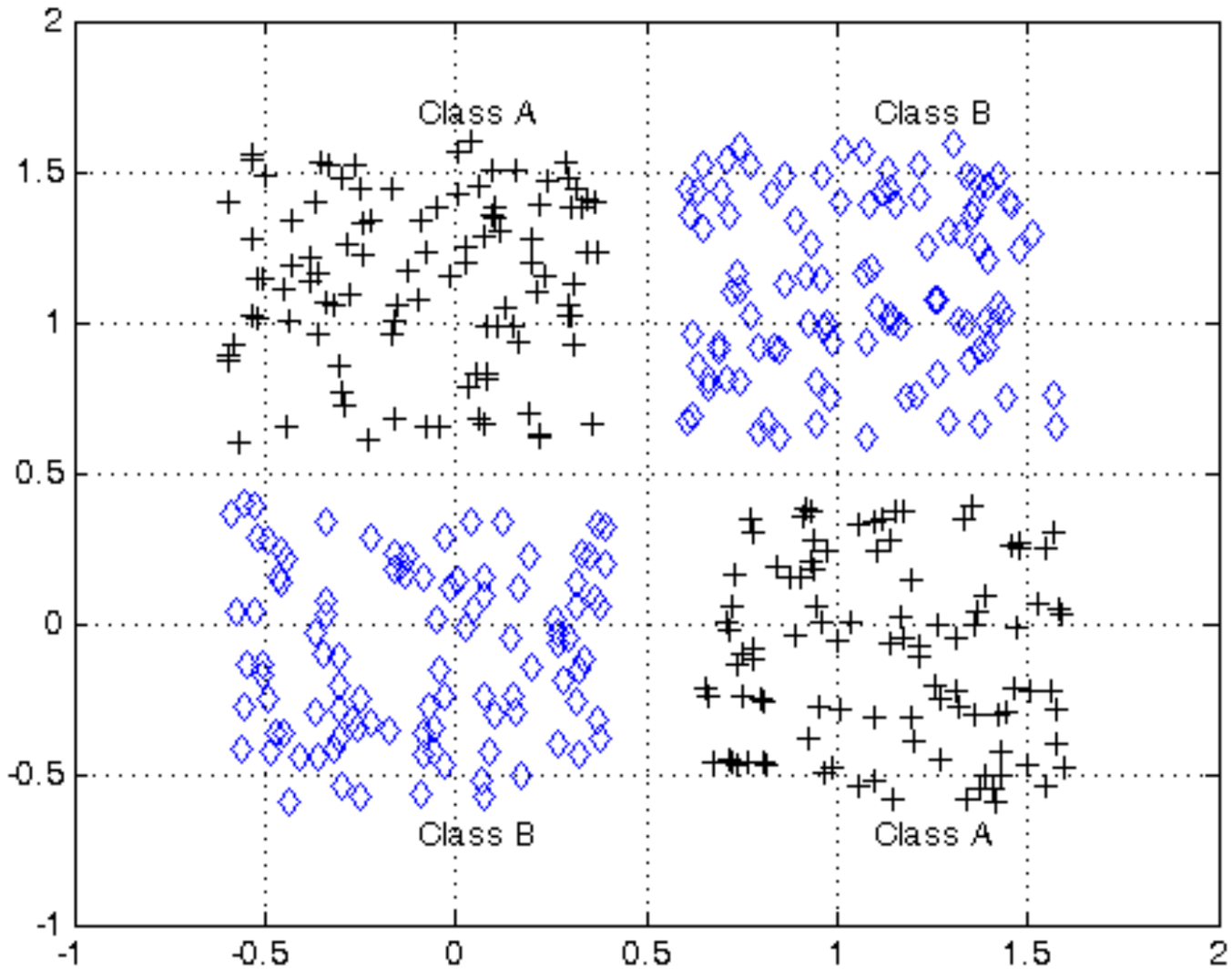


- $z = z(x) = a^*X_1 + b^*X_2 + c^*X_3 + d$

- $\hat{Y} = \frac{1}{1 + e^{-(a^*X_1 + b^*X_2 + c^*X_3 + d)}}$

Demo

XOR-problem



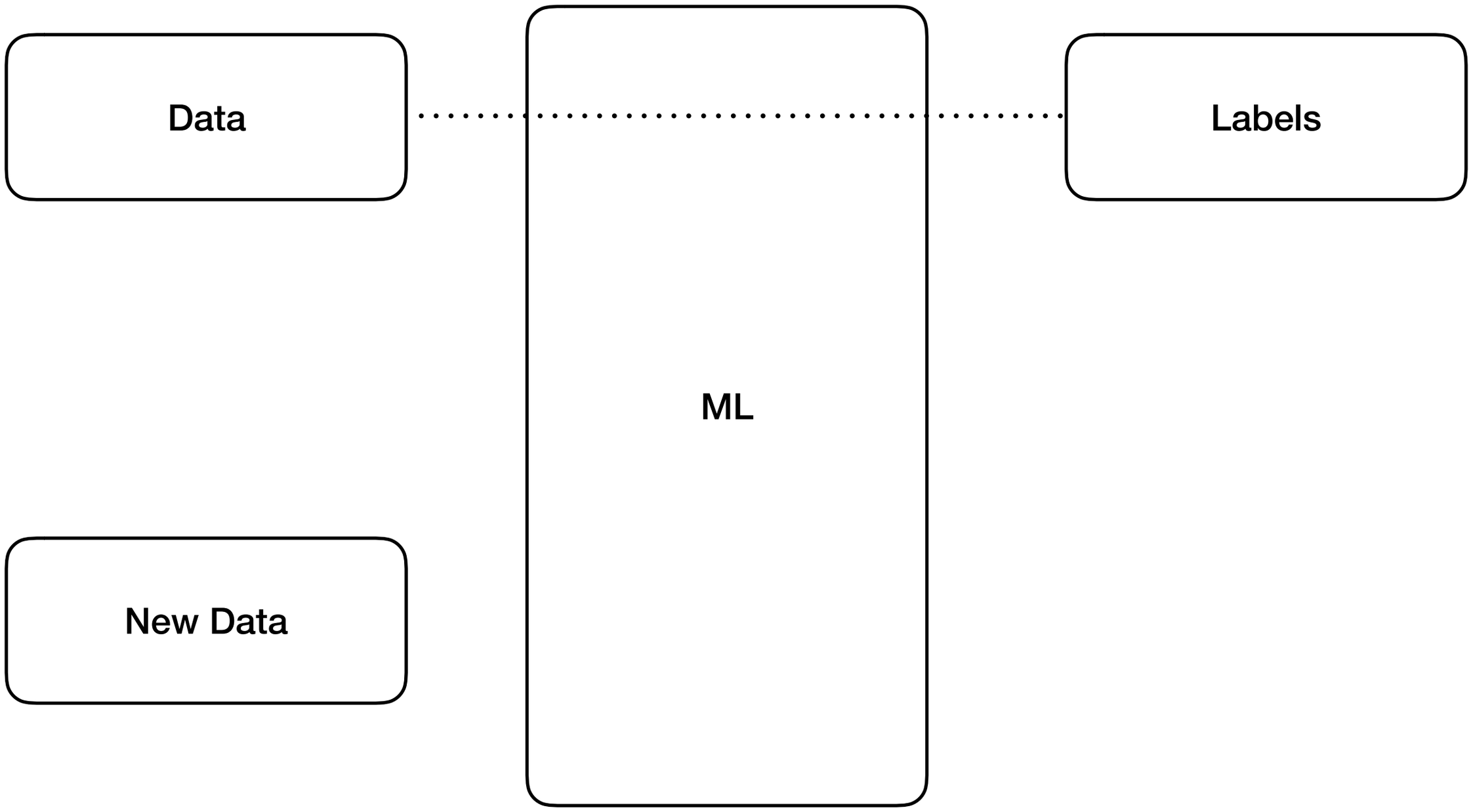
WorkFlow

WorkFlow

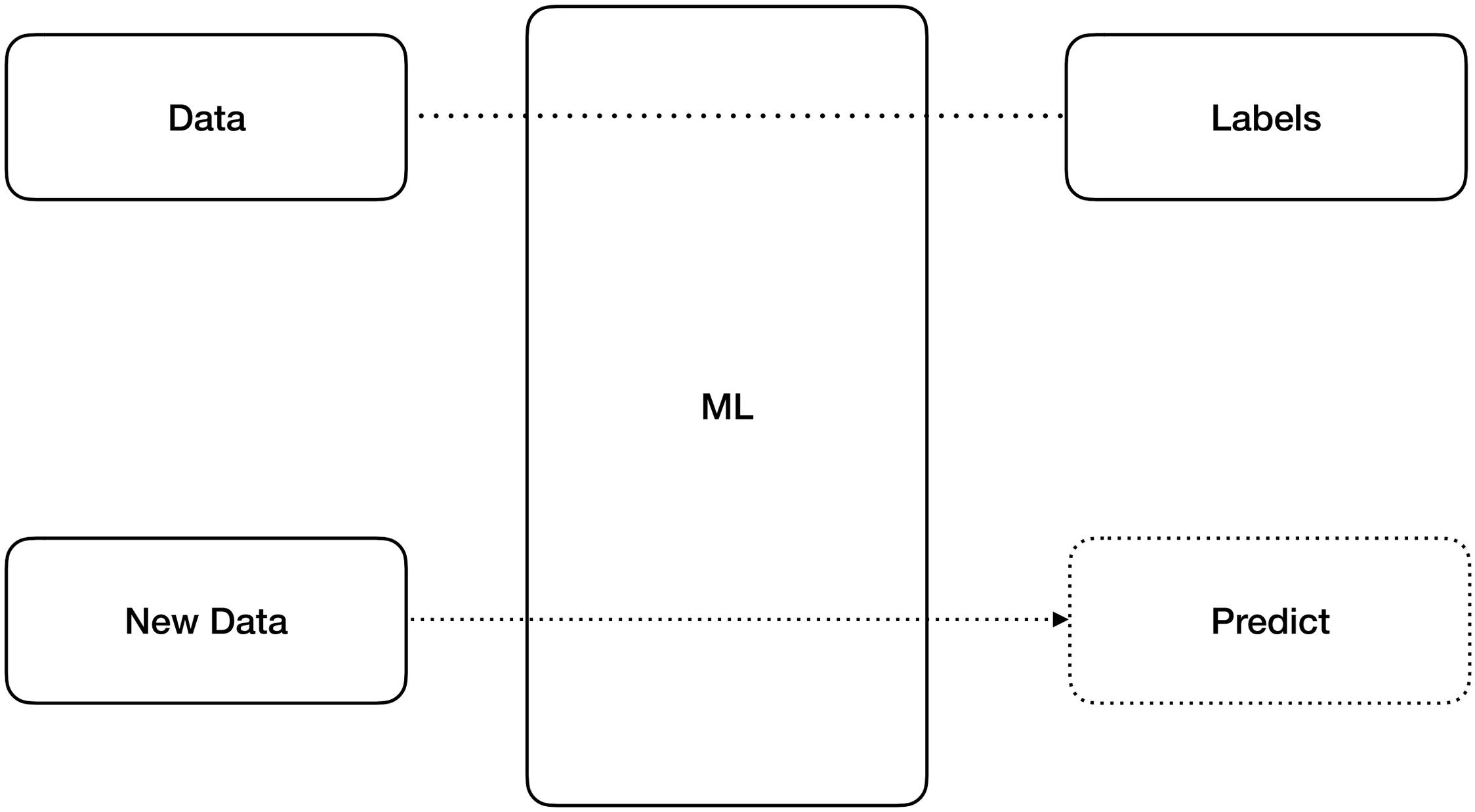
WorkFlow

WorkFlow

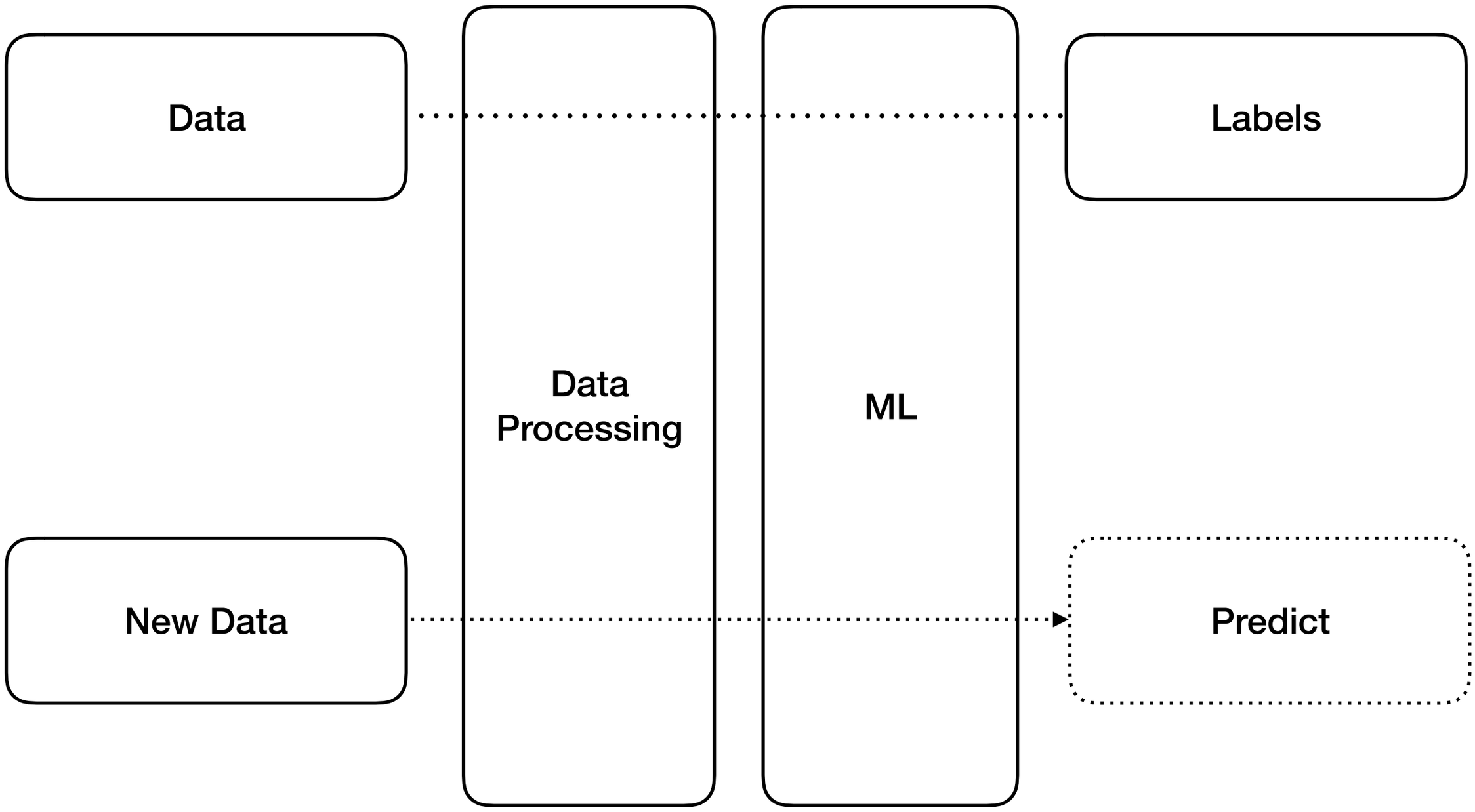
WorkFlow



WorkFlow



WorkFlow



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**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

All values of features are real numbers

But how to work with categorical non-numeric values?

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)	Type of apartment	Price (thous. $)
70	2	40	loft	100
130	3	15	business	160
40	1	5	studio	83
55	1	25	studio	**???**

Example: Apartment price

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

Categorical values

- Simple parameter coding:

Type of apartment
loft
business
studio
studio

Categorical values

- Simple parameter coding:
  1 - loft, 2 - business, 3 - studio

Type of apartment
loft
business
studio
studio

Categorical values

- Simple parameter coding:
  1 - loft, 2 - business, 3 - studio

Type of apartment
1
2
3
3

Categorical values

- Simple parameter coding:
  1 - loft, 2 - business, 3 - studio

Type of apartment
loft
business
studio
studio

Categorical values

- Simple parameter coding:
  1 - loft, 2 - business, 3 - studio

- One-hot-encoding (OHN)

Type of apartment
loft
business
studio
studio

Categorical values

- Simple parameter coding:
  1 - loft, 2 - business, 3 - studio

Type of apartment
loft
business
studio
studio

- One-hot-encoding (OHN)

type_loft	type_business	type_studio
1	0	0
0	1	0
0	0	1
0	0	1

Categorical values

- Simple parameter coding:
  1 - loft, 2 - business, 3 - studio

- One-hot-encoding (OHN)

type_loft	type_business	type_studio	Type of apartment
1	0	0	loft
0	1	0	business
0	0	1	studio
0	0	1	studio

Categorical values

- Simple parameter coding:
  1 - loft, 2 - business, 3 - studio

- One-hot-encoding (OHN)

type_loft	type_business	type_studio		Type of apartment
1	0	0	⟷	loft
0	1	0	⟷	business
0	0	1	⟷	studio
0	0	1	⟷	studio

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	**???**

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

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

$$\hat{Y} = f(X1, X2, X3)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d$$

Machine Learning Task:

find: *a, b, c, d*

and minimization of error

Example: Apartment price

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

Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	type_loft	type_business	type_studio	Price (thous. $)
70	2	40	1	0	0	100
130	3	15	0	1	0	160
40	1	5	0	0	1	83
55	1	25	0	0	1	???

Example: Apartment price

Total area (sq. m.)	Number of bedrooms	Age of the house (years)	type_loft	type_business	type_studio	Price (thous. $)
70	2	40	1	0	0	100
130	3	15	0	1	0	160
40	1	5	0	0	1	83
55	1	25	0	0	1	???

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

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

$$\hat{Y} = f(X1, X2, X3, X4, X5, X6)$$

$$\hat{Y} = a*X1 + b*X2 + c*X3 + d*X4 + e*X5 + f*X6 + j$$

Demo