Introduction

Probability of default
ROC curve and AUC

Szkoła Główna Handlowa w Warszawie

Warsaw, March 2023

Logistic regression

Real-life examples:

- Predict whether or not a customer will default on a loan
- Predict whether or not a patient will have a heart attack
- Predict whether or not an email is a spam
- Predict whether or not a student will pass/fail an exam
- Predict whether the patient has covid

Classical formula:

$$p_i = \frac{1}{1 + \ e^{-(\beta_1 x_{1,i} + \beta_2 x_{2,i} + \ \ldots \beta_k x_{k,i})}} \in (0,1)$$

Performance metrics

Prediction	11.2%	21.3%	32.1%	44.3%	52.4%	61.7%	70.9%	81.9%	94.0%	99.9%
Actual output	0	0	1	0	0	1	1	0	1	0

How to measure whether the model is „good"?

Predictive power

Checks whether the model predictions accurately reflects the true (unknown) probabilities (~for **how many units** the actual output will be „one")

- binomial test
- Hosmer-Lemeshow test
- Spiegelhalter test
- Jeffrey's test

Discriminatory power

Check whether the model is able to „catch" cases with highest (lowest) probabilities (~for **which units** the actual output will be „one")

- ROC curve
- AUC statistic
- c-index

Simple example

Threshold: 50%

0 | 1 - model's final prediction

Predictions	11.2%	21.3%	32.1%	44.3%	52.4%	61.7%	70.9%	81.9%	94.0%	99.9%
Actual output	0	0	1	0	0	1	1	0	1	0

Confusion matrix:

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	Positive (1)	True positive	False negative
	Negative (0)	False positive	True negative

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	Positive (1)	3	1
	Negative (0)	3	3

Simple example

Threshold: 70%

0 | 1 - model's final prediction

Predictions	11.2%	21.3%	32.1%	44.3%	52.4%	61.7%	70.9%	81.9%	94.0%	99.9%
Actual output	0	0	1	0	0	1	1	0	1	0

Confusion matrix:

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	Positive (1)	True positive	False negative
	Negative (0)	False positive	True negative

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	Positive (1)	2	2
	Negative (0)	2	4

Simple example

Confusion matrix:

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	[P]ositive (1)	True positive (TP)	False negative (FN)
	[N]egative (0)	False positive (FP)	True negative (TN)

Covid example:

		Test prediction	
		Positive (1)	Negative (0)
Actual Condition	[P]ositive (1)	Patients with COVID for which test result was positive	Patients with COVID for which test result was negative
	[N]egative (0)	Patients without COVID for which test result was positive	Patients without COVID for which test result was negative

Metrics:

- Accuracy: $\frac{TP+TN}{ALL}$ — How many times the prediction was correct?
- Recall: $\frac{TP}{P}$ — Among all patients with COVID (P), how many of them received positive test results?
- Precision: $\frac{TP}{TP+FP}$ — Among all patients with positive test results (TP + FP), how many of them have actually COVID?
- Specificity: $\frac{TN}{N}$ — Among all patients without COVID, how many of them didn't actually have it?

Simple example

Threshold: 70%

$\underbrace{\hspace{6cm}}_{}$ 0 $\quad$ 1 - model's final prediction

Predictions	11.2%	21.3%	32.1%	44.3%	52.4%	61.7%	70.9%	81.9%	94.0%	99.9%
Actual output	0	0	1	0	0	1	1	0	1	0

Confusion matrix:

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	Positive (1)	True positive	False negative
	Negative (0)	False positive	True negative

Metrics:

- Recall: $\frac{TP}{P}$
- Specificity: $\frac{TN}{N}$

Simple example

Actual value	Prediction
0	11.2%
0	21.3%
0	32.1%
0	44.3%
1	52.4%
0	61.7%
1	70.9%
0	81.9%
1	94.0%
0	99.9%

1

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	Positive (1)	True positive	False negative
	Negative (0)	False positive	True negative

Metrics:

- Recall: $\frac{TP}{P}$
- False positive rate: $\frac{FP}{N}$

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	Positive (1)	3	0
	Negative (0)	7	0

Metrics:

- Recall: $\frac{3}{3}$
- False positive rate: $\frac{7}{7}$

Simple example

Actual value	Prediction
0	11.2%
0	21.3%
0	32.1%
0	44.3%
1	52.4%
0	61.7%
1	70.9%
0	81.9%
1	94.0%
0	99.9%

0

1

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	Positive (1)	True positive	False negative
	Negative (0)	False positive	True negative

Metrics:

- Recall: $\frac{TP}{P}$
- False positive rate: $\frac{FP}{N}$

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	Positive (1)	3	0
	Negative (0)	6	1

Metrics:

- Recall: $\frac{3}{3}$
- False positive rate: $\frac{6}{7}$

Simple example

Actual value	Prediction
0	11.2%
0	21.3%
0	32.1%
0	44.3%
1	52.4%
0	61.7%
1	70.9%
0	81.9%
1	94.0%
0	99.9%

0

1

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	Positive (1)	True positive	False negative
	Negative (0)	False positive	True negative

Metrics:

- Recall: $\frac{TP}{P}$
- False positive rate: $\frac{FP}{N}$

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	Positive (1)	3	0
	Negative (0)	5	2

Metrics:

- Recall: $\frac{3}{3}$
- False positive rate: $\frac{5}{7}$

Simple example

Actual value	Prediction
0	11.2%
0	21.3%
0	32.1%
0	44.3%
1	52.4%
0	61.7%
1	70.9%
0	81.9%
1	94.0%
0	99.9%

Predictions 11.2%–61.7%: 0

Predictions 70.9%–99.9%: 1

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	Positive (1)	True positive	False negative
	Negative (0)	False positive	True negative

Metrics:

- Recall: $\frac{TP}{P}$
- False positive rate: $\frac{FP}{N}$

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	Positive (1)	2	1
	Negative (0)	2	5

Metrics:

- Recall: $\frac{2}{3}$
- False positive rate: $\frac{2}{7}$

Simple example

Actual value	Prediction
0	11.2%
0	21.3%
0	32.1%
0	44.3%
1	52.4%
0	61.7%
1	70.9%
0	81.9%
1	94.0%
0	99.9%

$\Big\} \; 0$

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	Positive (1)	True positive	False negative
	Negative (0)	False positive	True negative

		Predicted condition	
		Positive (1)	Negative (0)
Actual condition	Positive (1)	0	3
	Negative (0)	0	7

Metrics:

- Recall: $\frac{TP}{P}$
- False positive rate: $\frac{FP}{N}$

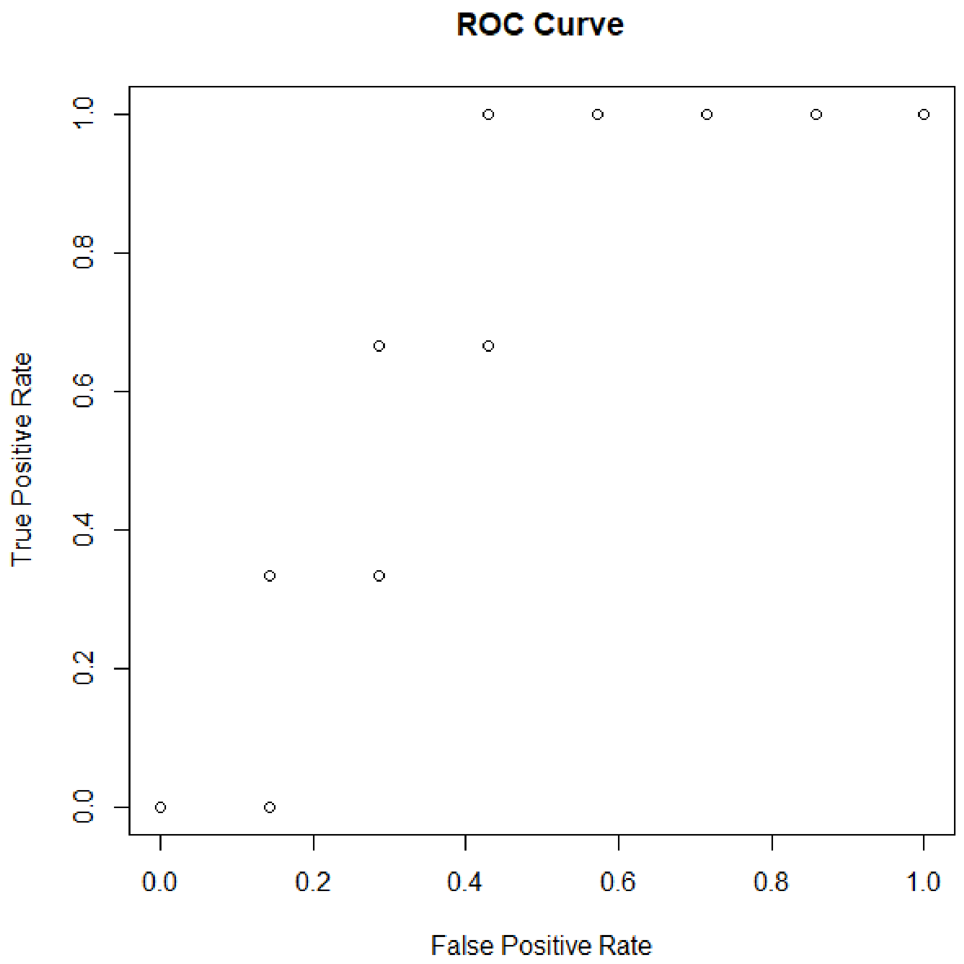
Metrics:

- Recall: $\frac{0}{3}$
- False positive rate: $\frac{0}{7}$

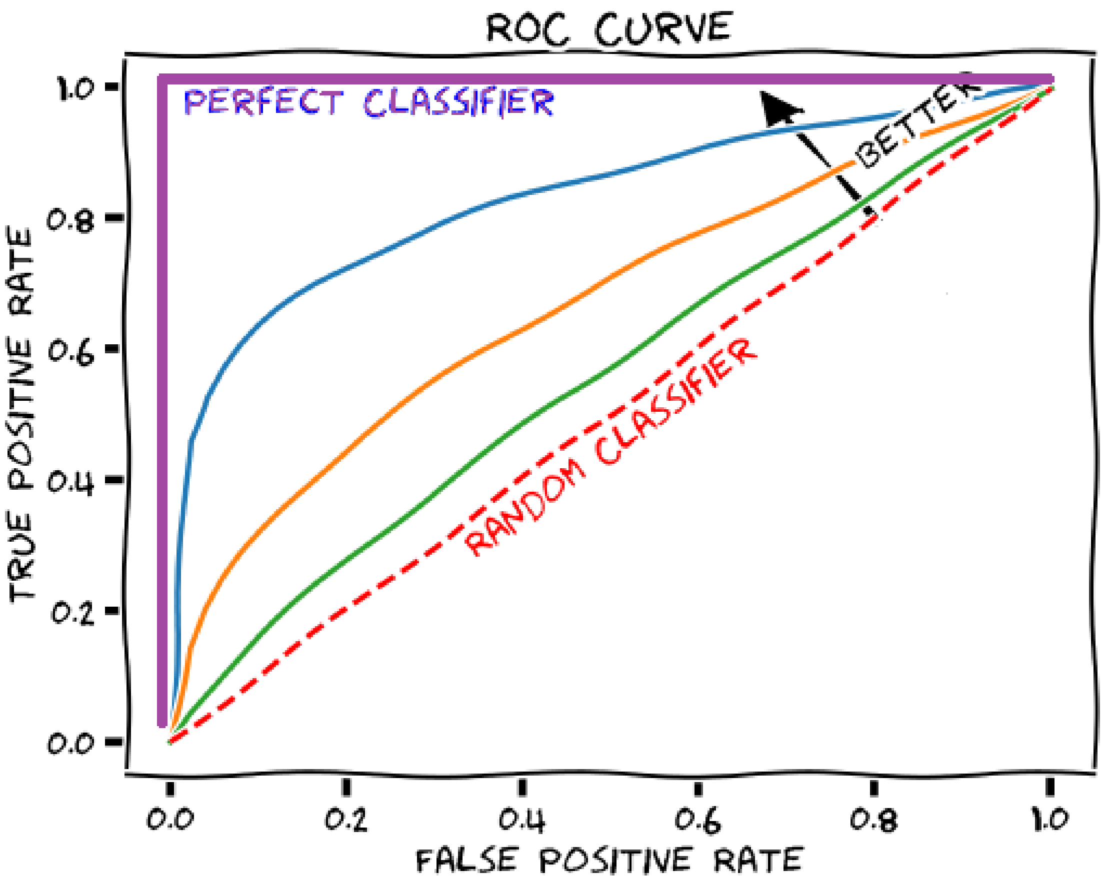
ROC curve

Actual value	Prediction
0	11.2%
0	21.3%
0	32.1%
0	44.3%
1	52.4%
0	61.7%
1	70.9%
0	81.9%
1	94.0%
0	99.9%

Recall	False Positive Rate
3/3	7/7
3/3	6/7
3/3	5/7
3/3	4/7
3/3	3/7
2/3	3/7
2/3	2/7
1/3	2/7
1/3	1/7
0/3	1/7
0/3	0/7



AUC's properties



0 < AUC < 1

0.5 < AUC:	model better than random sampling
0.5 = AUC:	model does not have any discriminatory power (is „as good" as random sampling)
AUC < 0.5:	it is better not to have this model (or is it?)

ROC – **R**eceiver **O**perating **C**haracteristic
AUC – **A**rea **U**nder **C**urve

AUC's properties

- 0 ≤ AUC ≤ 1 by definition
- AUC of perfect model = 1
- AUC of worst model = 0
- AUC of „random" ~ 0.5
- AUC($\hat{y}$, y) = 1 - AUC(1 - $\hat{y}$, y)
- Monotonic transformation of $\hat{y}$ does not change AUC, e.g.

  AUC($\hat{y}$, y) = AUC($\ln(2 \ * \hat{y} + 20)$, y)
- Even AUC ~ 1 does not ensure the proper performance
- Has the probability interpretation
- Can be perceived as special case of c-index statistics

0.5 < AUC: model better than random sampling

0.5 = AUC: model does not have any discriminatory power (is „as good" as random sampling)

AUC < 0.5: it is better not to have this model (or is it?)

$\hat{y}$ - Vector of probabilites* (e.g. [0.1, 0.15, 0.54, 0.39])

y - Vector of actual binary output (e.g. [1, 0, 1, 1])

AUC – probabilistic interpretation

Predictions	$\widehat{p}_i$	11.2%	21.3%	32.1%	44.3%	52.5%	60%
Actual output	y_i	0	0	1	0	1	0

What is a probability that $P(\,\widehat{p}_i > \widehat{p}_j \,|(\,y_i = 1\ AND\ y_j = 0\,))$? P(...) = 5 / 8

i	j	$\widehat{p}_i > \widehat{p}_j$	i	j	$\widehat{p}_i > \widehat{p}_j$
1	3	TRUE	4	3	FALSE
1	5	TRUE	4	5	TRUE
2	3	TRUE	6	3	FALSE
2	5	TRUE	6	5	FALSE

$$AUC = P(\,\widehat{p}_i > \widehat{p}_j \,|(\,y_i = 1\ AND\ y_j = 0\,)) + \frac{1}{2} \cdot P(\,\widehat{p}_i = \widehat{p}_j \,|(\,y_i = 1\ AND\ y_j = 0\,))$$

AUC – special case of concordance index (c-index / Harrell's c-index)

Predictions	$\hat{p}_i$	11.2%	21.3%	32.1%	44.3%	52.5%	60%
Actual output	y_i	0.1	0.13	0.44	0.24	0.59	0.46

What is a probability that $P(\hat{p}_i > \hat{p}_j \,|(y_i > y_j))$?

$$\text{c-index} = P(\hat{p}_i > \hat{p}_j \mid y_i > y_j)) + \frac{1}{2} \cdot P(\hat{p}_i = \hat{p}_j \,|(y_i > y_j))$$

https://statisticaloddsandends.wordpress.com/2019/10/26/what-is-harrells-c-index/

Questions

Introduction

Sampling
Development and validation samples
Out-of-sample and out-of-time
Cross-validation

Szkoła Główna Handlowa w Warszawie

Warsaw, March 2023

Introduction to Probability of Default modeling

EL = EAD X PD X LGD

EL – Expected Loss

EAD – Exposure at Default

Expected outstandings at the moment of default

PD – Probability of Default

Probability of customer's default within 12 months

LGD – Loss given Default

Ultimate loss as a % of Exposure at the moment of default

Types of sampling

Sampling

Random sampling

Subset of individuals (a sample) chosen from a larger set (a population) in which a subset of individuals are chosen randomly, all with the same probabilityadf

Undersampling

Undersampling is a technique to balance uneven datasets by keeping all of the data in the minority class and decreasing the size of the majority class.

Oversampling

Random oversampling involves supplementing the training data with multiple copies of some of the minority classes.

Systematic sampling

Systematic sampling is a probability sampling method where researchers select members of the population at a regular interval

Stratified sampling

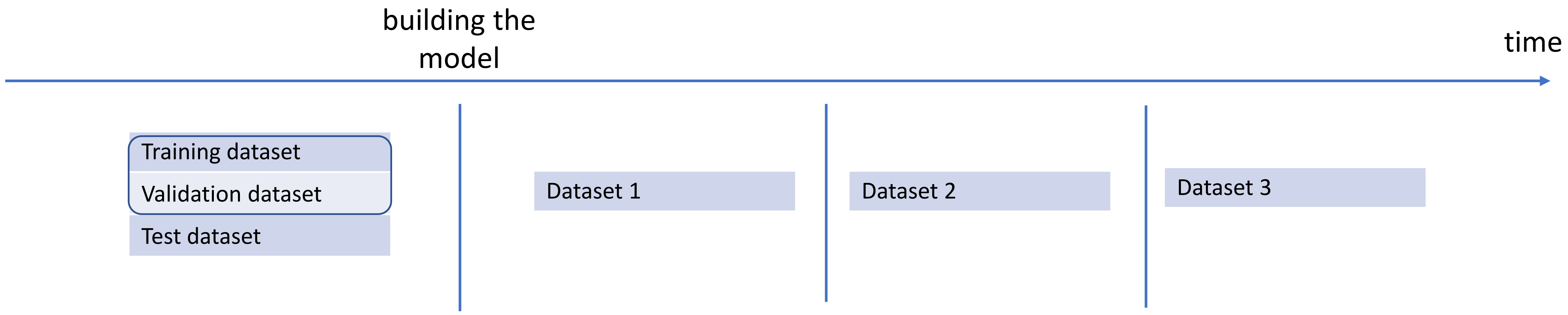
Stratification is the process of dividing members of the population into homogeneous subgroups before sampling. The strata should define a partition of the population. That is, it should be collectively exhaustive and mutually exclusive: every element in the population must be assigned to one and only one stratum.

Cluster sampling

Cluster sampling is a probability sampling method in which you divide a population into clusters, such as districts or schools, and then randomly select some of these clusters as your sample.

Splitting dataset

Splitting dataset



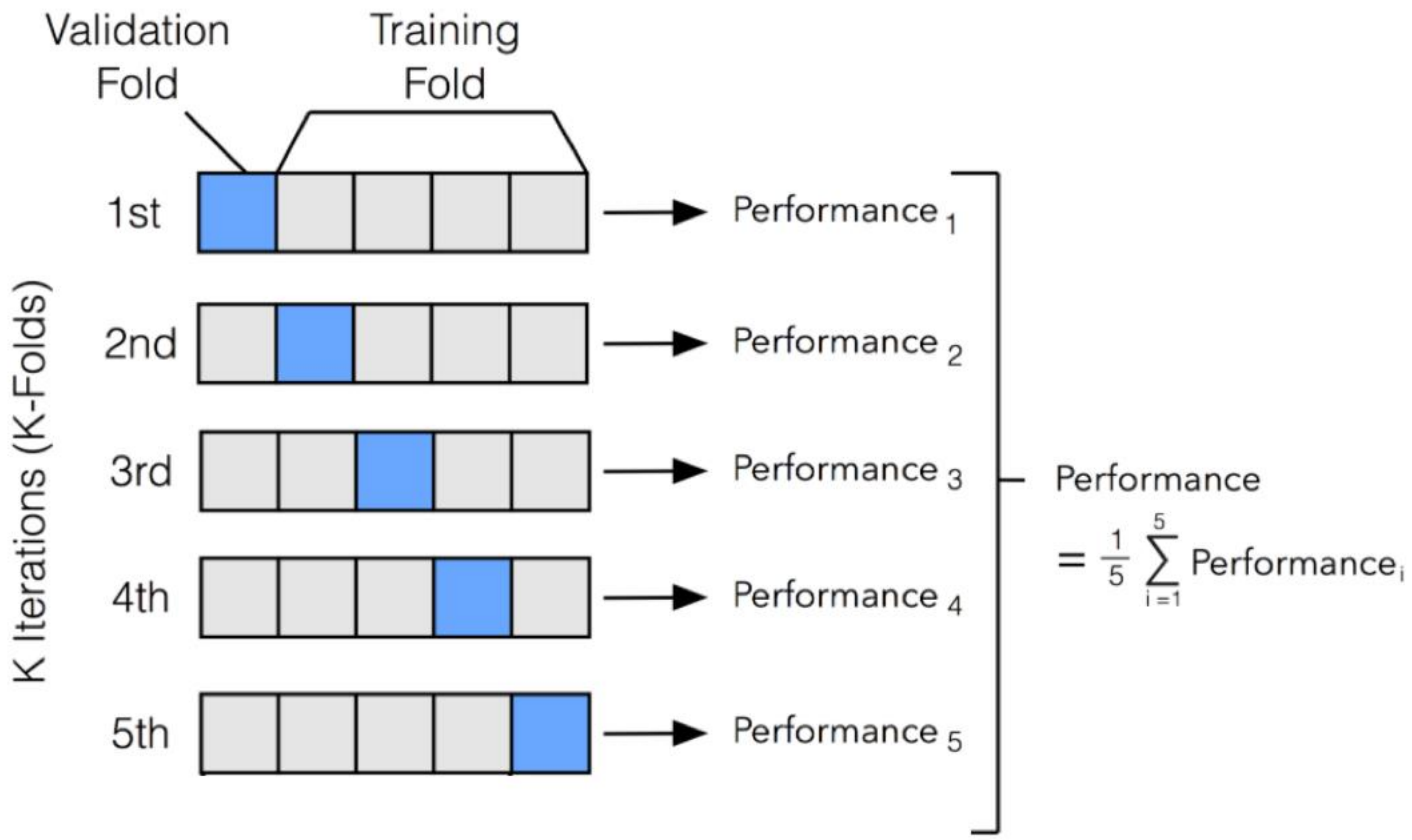
Development sample = training dataset + validation dataset

~ model development perspective

Validation samples = Dataset 1, Dataset 2, Dataset 3, Test dataset
Out-of-time = Dataset 1, Dataset 2, Dataset 3, ...

~ model validation perspective

Cross-validation



Questions