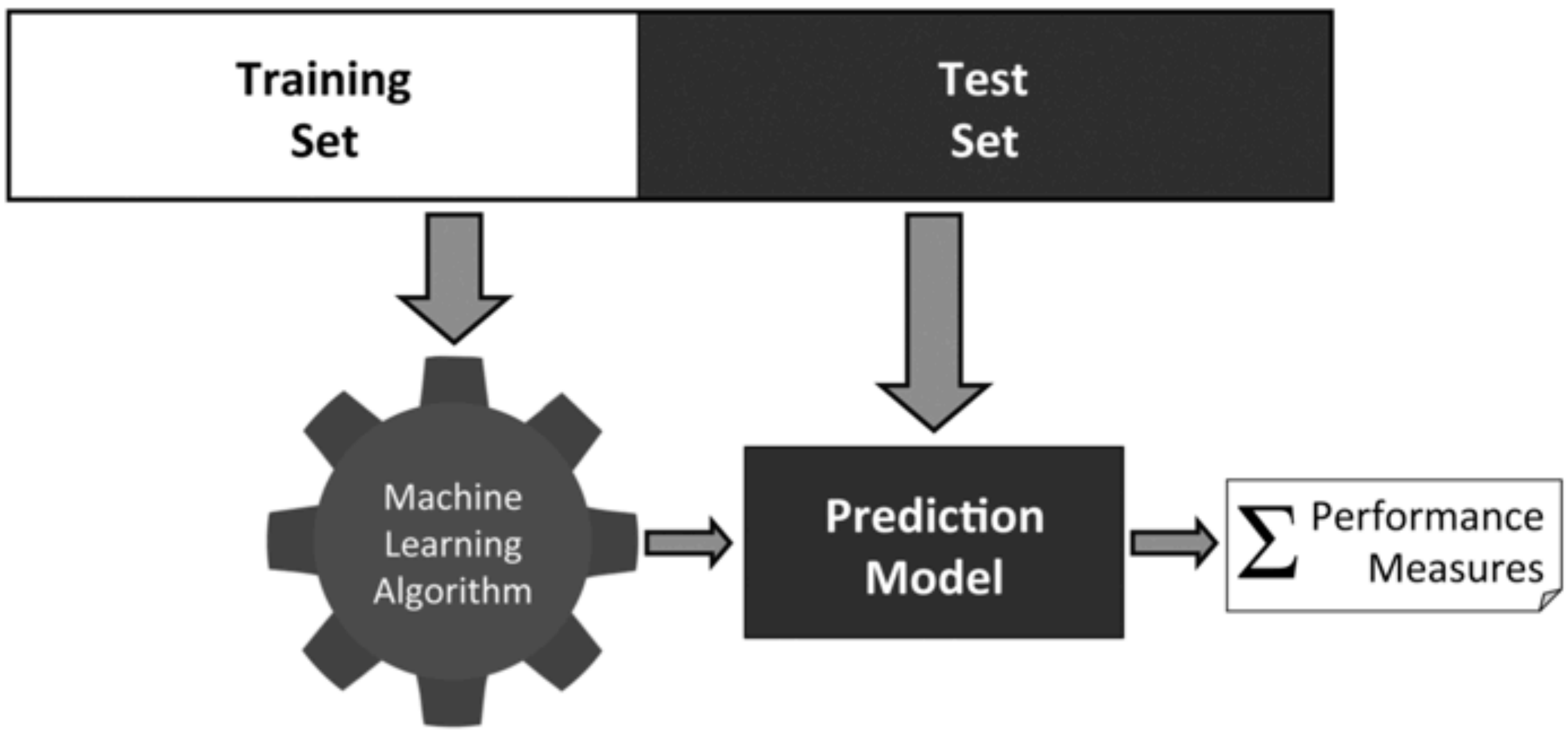
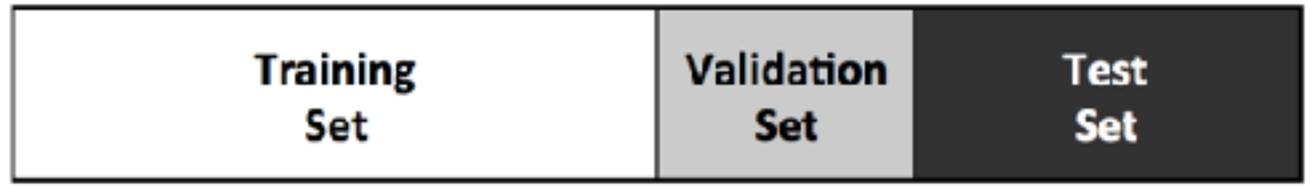
高等資料探勘與巨量資料處理

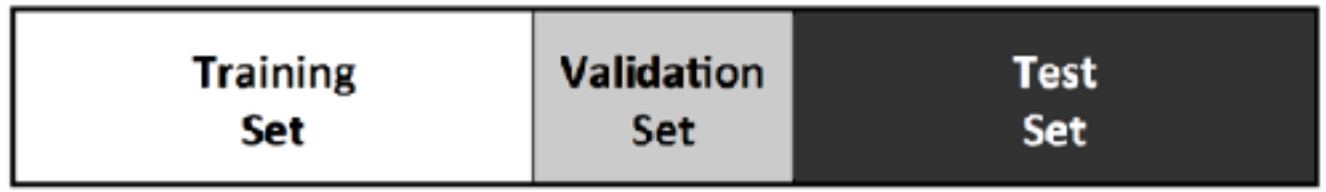
Performance Evaluation


Training
Set
Test
Set
Machine
Learning
Algorithm
Prediction
Model
Performance
Measures

Common Evaluation Experiment Setting



(a) A 50:20:30 split



(b) A 40:20:40 split

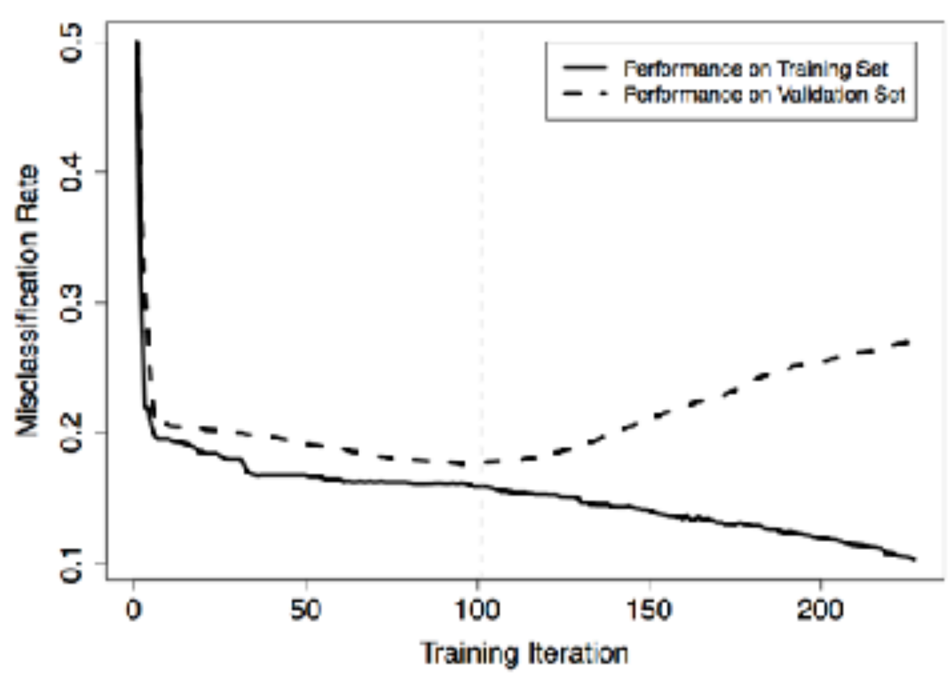
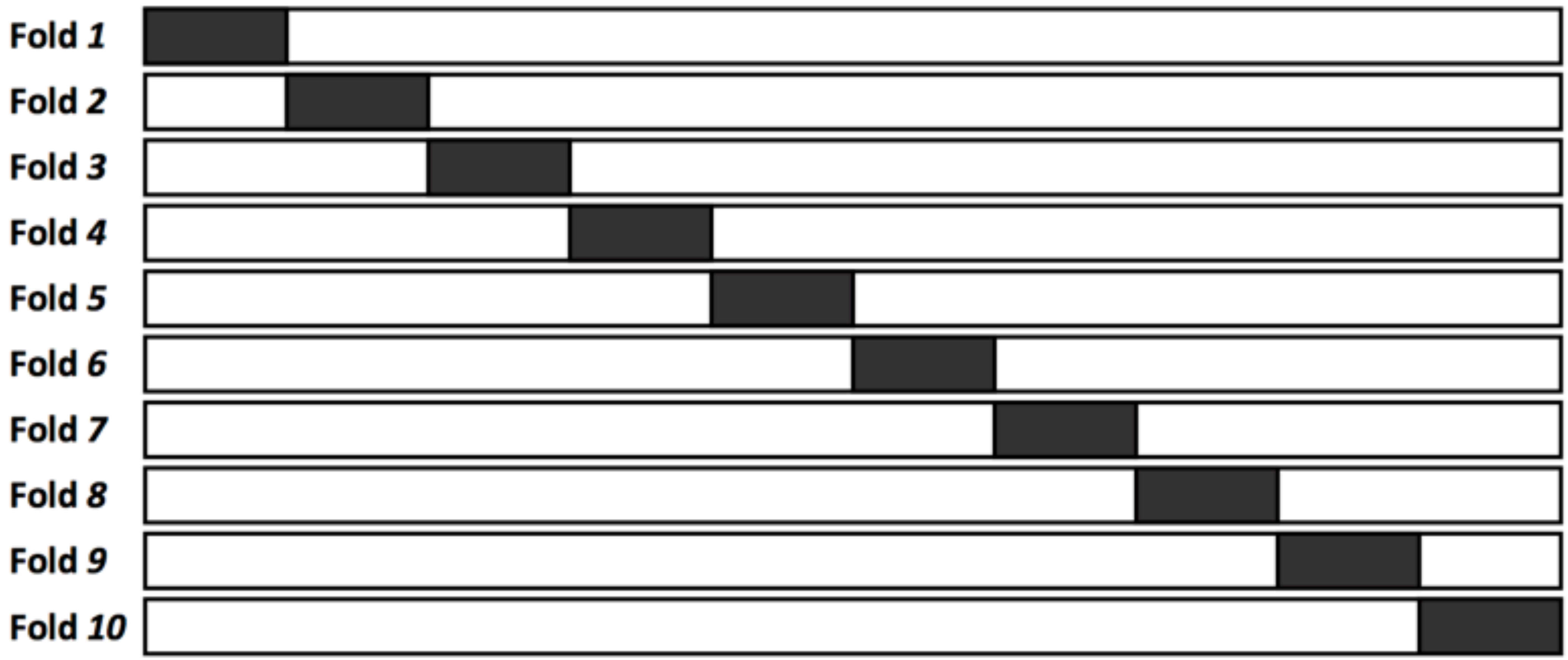


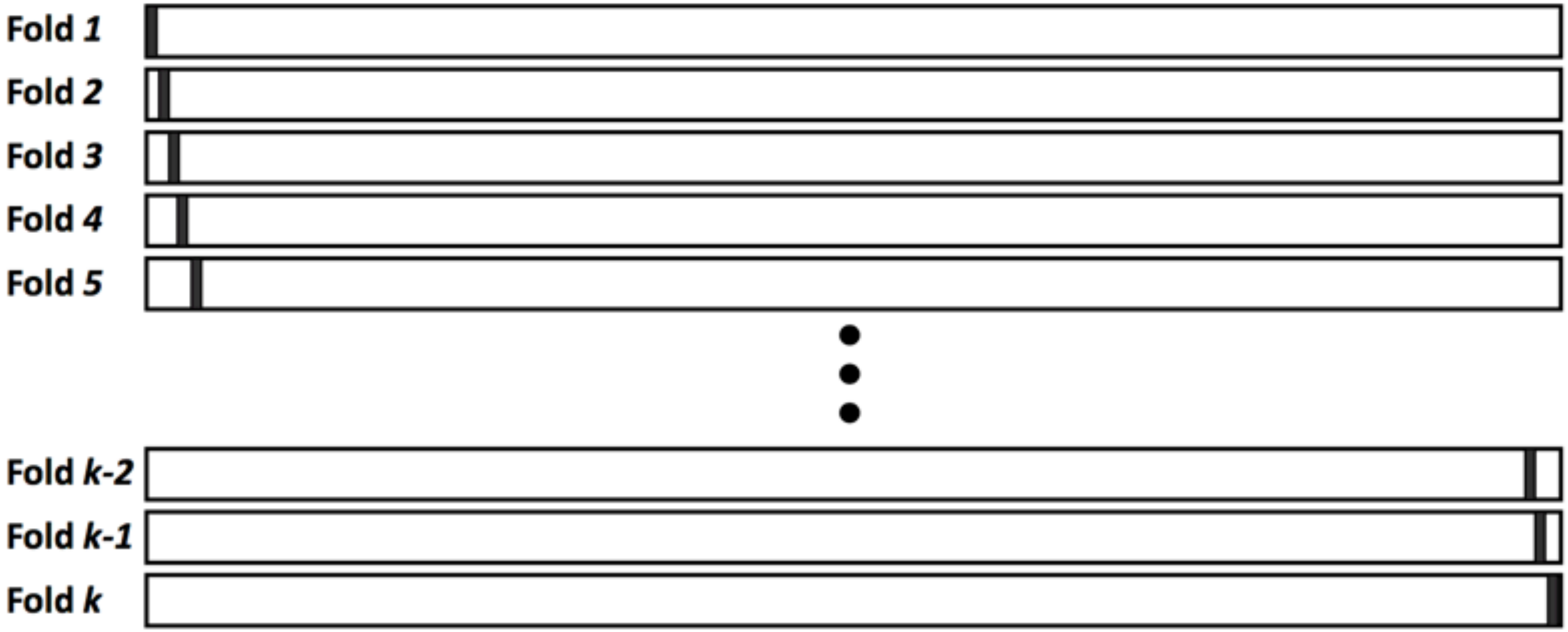
Figure: Using a validation set to avoid overfitting in iterative machine learning algorithms.

k-fold cross validation



The division of data during the *k*-fold cross validation process. Black rectangles indicate test data, and white spaces indicate training data.

leave-one-out cross validation



Common Performance Measure for Classification

Table: A sample test set with model predictions.

ID	Target	Pred.	Outcome	ID	Target	Pred.	Outcome
1	spam	ham	FN	11	ham	ham	TN
2	spam	ham	FN	12	spam	ham	FN
3	ham	ham	TN	13	ham	ham	TN
4	spam	spam	TP	14	ham	ham	TN
5	ham	ham	TN	15	ham	ham	TN
6	spam	spam	TP	16	ham	ham	TN
7	ham	ham	TN	17	ham	spam	FP
8	spam	spam	TP	18	spam	spam	TP
9	spam	spam	TP	19	ham	ham	TN
10	spam	spam	TP	20	ham	spam	FP

For binary prediction problems there are 4 possible outcomes:

1. True Positive (TP)
2. True Negative (TN)
3. False Positive (FP)
4. False Negative (FN)

		Prediction	
		'spam'	***'ham'***
Target	***'spam'***	6	3
	'ham'	2	9

Table: The structure of a confusion matrix.

		Prediction	
		positive	**negative**
Target	**positive**	*TP*	*FN*
	negative	*FP*	*TN*

$$\text{classification accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

$$\text{classification accuracy} = \frac{(6 + 9)}{(6 + 9 + 2 + 3)} = 0.75$$

		Prediction	
		positive	**negative**
Target	**positive**	*TP*	*FN*
	negative	*FP*	*TN*

		Prediction	
		'spam'	***'ham'***
Target	***'spam'***	6	3
	'ham'	2	9

$$\text{precision} = \frac{TP}{(TP + FP)} = \frac{6}{(6+2)} = 0.75$$

$$\text{recall} = \frac{TP}{(TP + FN)} = \frac{6}{(6+3)} = 0.667$$

$$\mathrm{F_1\text{-}measure} = 2 \times \frac{(\mathrm{precision} \times \mathrm{recall})}{(\mathrm{precision} + \mathrm{recall})}$$

$$\begin{aligned}\mathrm{F_1\text{-}measure} &= 2 \times \frac{\left(\frac{6}{(6+2)} \times \frac{6}{(6+3)}\right)}{\left(\frac{6}{(6+2)} + \frac{6}{(6+3)}\right)} \\ &= 0.706\end{aligned}$$

$$\mathrm{precision} = \frac{TP}{(TP + FP)} = \frac{6}{(6+2)} = 0.75$$

$$\mathrm{recall} = \frac{TP}{(TP + FN)} = \frac{6}{(6+3)} = 0.667$$

		Prediction	
		positive	negative
Target	**positive**	*TP*	*FN*
	negative	*FP*	*TN*

		Prediction	
		'spam'	*'ham'*
Target	*'spam'*	6	3
	'ham'	2	9

$$\mathrm{TPR} = \frac{TP}{(TP + FN)}$$

$$\mathrm{TNR} = \frac{TN}{(TN + FP)}$$

$$\mathrm{FPR} = \frac{FP}{(TN + FP)}$$

$$\mathrm{FNR} = \frac{FN}{(TP + FN)}$$

$$\mathrm{TPR} = \frac{6}{(6+3)} = 0.667$$

$$\mathrm{TNR} = \frac{9}{(9+2)} = 0.818$$

$$\mathrm{FPR} = \frac{2}{(9+2)} = 0.182$$

$$\mathrm{FNR} = \frac{3}{(6+3)} = 0.333$$

TPR+FNR=1

TNR+FPR=1

Let's compare the following two models for the same problem

		Prediction	
		'non-churn'	*'churn'*
Target	*'non-churn'*	90	0
	'churn'	9	1

		Prediction	
		'non-churn'	*'churn'*
Target	*'non-churn'*	70	20
	'churn'	2	8

Use Simple Accuracy for a Reference?

$$\text{average class accuracy} = \frac{1}{|levels(t)|} \sum_{l \in levels(t)} \text{recall}_l$$

Harmonic Mean

$$\text{average class accuracy}_{HM} = \frac{1}{\frac{1}{|\mathit{levels}(t)|} \sum_{l \in \mathit{levels}(t)} \frac{1}{\text{recall}_l}}$$

		Prediction	
		'non-churn'	*'churn'*
Target	*'non-churn'*	90	0
	'churn'	9	1

$$\frac{1}{\frac{1}{2}\left(\frac{1}{1.0} + \frac{1}{0.1}\right)} = \frac{1}{5.5} = 18.2\%$$

		Prediction	
		'non-churn'	*'churn'*
Target	*'non-churn'*	70	20
	'churn'	2	8

$$\frac{1}{\frac{1}{2}\left(\frac{1}{0.778} + \frac{1}{0.800}\right)} = \frac{1}{1.268} = 78.873\%$$

What about the outcomes are not equally important?

Which one is better ?

Model A

		Prediction 'good'	Prediction 'bad'
Target	'good'	57	3
	'bad'	10	30

Model B

		Prediction 'good'	Prediction 'bad'
Target	'good'	43	17
	'bad'	3	37

Exercise: Please compute Average Class Accuracy in terms of Harmonic Mean ?

Employ a profit matrix:

		Prediction positive	Prediction negative
Target	positive	TP_{Profit}	FN_{Profit}
	negative	FP_{Profit}	TN_{Profit}

The **profit matrix** for the pay-day loan credit scoring problem.

		Prediction *'good'*	Prediction *'bad'*
Target	*'good'*	140	−140
	'bad'	−700	0

Model A

		Prediction *'good'*	Prediction *'bad'*
Target	*'good'*	57	3
	'bad'	10	30

Model B

		Prediction *'good'*	Prediction *'bad'*
Target	*'good'*	43	17
	'bad'	3	37

		Prediction	
		'good'	*'bad'*
Target	*'good'*	57	3
	'bad'	10	30

		Prediction	
		'good'	*'bad'*
Target	*'good'*	7 980	−420
	'bad'	−7 000	0
	Profit		**560**

		Prediction	
		'good'	*'bad'*
Target	*'good'*	57	3
	'bad'	10	30

		Prediction	
		'good'	*'bad'*
Target	*'good'*	6 020	−2 380
	'bad'	−2 100	0
	Profit		**1 540**

Performance Measures: Multinomial Targets

ID	Target	Prediction	ID	Target	Prediction
1	durionis	fructosus	16	ficulneus	ficulneus
2	ficulneus	fructosus	17	ficulneus	ficulneus
3	fructosus	fructosus	18	fructosus	fructosus
4	ficulneus	ficulneus	19	durionis	durionis
5	durionis	durionis	20	fructosus	fructosus
6	pseudo.	pseudo.	21	fructosus	fructosus
7	durionis	fructosus	22	durionis	durionis
8	ficulneus	ficulneus	23	fructosus	fructosus
9	pseudo.	pseudo.	24	pseudo.	fructosus
10	pseudo.	fructosus	25	durionis	durionis
11	fructosus	fructosus	26	pseudo.	pseudo.
12	ficulneus	ficulneus	27	fructosus	fructosus
13	durionis	durionis	28	ficulneus	ficulneus
14	fructosus	fructosus	29	fructosus	fructosus
15	fructosus	ficulneus	30	fructosus	fructosus

Table: The structure of a confusion matrix for a multinomial predictior problem with l target levels.

		Prediction					Recall
		level1	*level2*	*level3*	$\cdots$	*levell*	
	level1	-	-	-		-	-
	level2	-	-	-		-	-
Target	*level3*	-	-	-		-	-
	$\vdots$				$\ddots$		$\vdots$
	levell	-	-	-		-	-
	Precision	-	-	-	$\cdots$	-	

$$\text{precision(l)} = \frac{TP(l)}{TP(l) + FP(l)}$$

$$\text{recall(l)} = \frac{TP(l)}{TP(l) + FN(l)}$$

		Prediction				Recall
		'durionis'	***'ficulneus'***	***'fructosus'***	***'pseudo.'***	
Target	***'durionis'***	5	0	2	0	0.714
	'ficulneus'	0	6	1	0	0.857
	'fructosus'	0	1	10	0	0.909
	'pseudo.'	0	0	2	3	0.600
	Precision	1.000	0.857	0.667	1.000	

$$\frac{1}{\frac{1}{4}\left(\frac{1}{0.714} + \frac{1}{0.857} + \frac{1}{0.909} + \frac{1}{0.600}\right)} = \frac{1}{1.333} = 75.000\%$$