



HOKKAIDO  
UNIVERSITY

# 講義「人工知能」

## 第5回 ニューラルネットワーク 2

### 学習結果の評価

北海道大学大学院情報科学研究所  
情報理工学部門 複合情報工学分野  
調和系工学研究室 准教授 山下倫央

<http://harmo-lab.jp>

[tomohisa@ist.hokudai.ac.jp](mailto:tomohisa@ist.hokudai.ac.jp)

2024年4月23日(火)

## ❖ モデルの評価

### ■ 2クラス分類

- 正解率 : Accuracy
- 適合率 : Precision
- 再現率 (感度) : Recall, True Positive Rate(TPR)
- F値 : F-measure

### ■ 多クラス分類

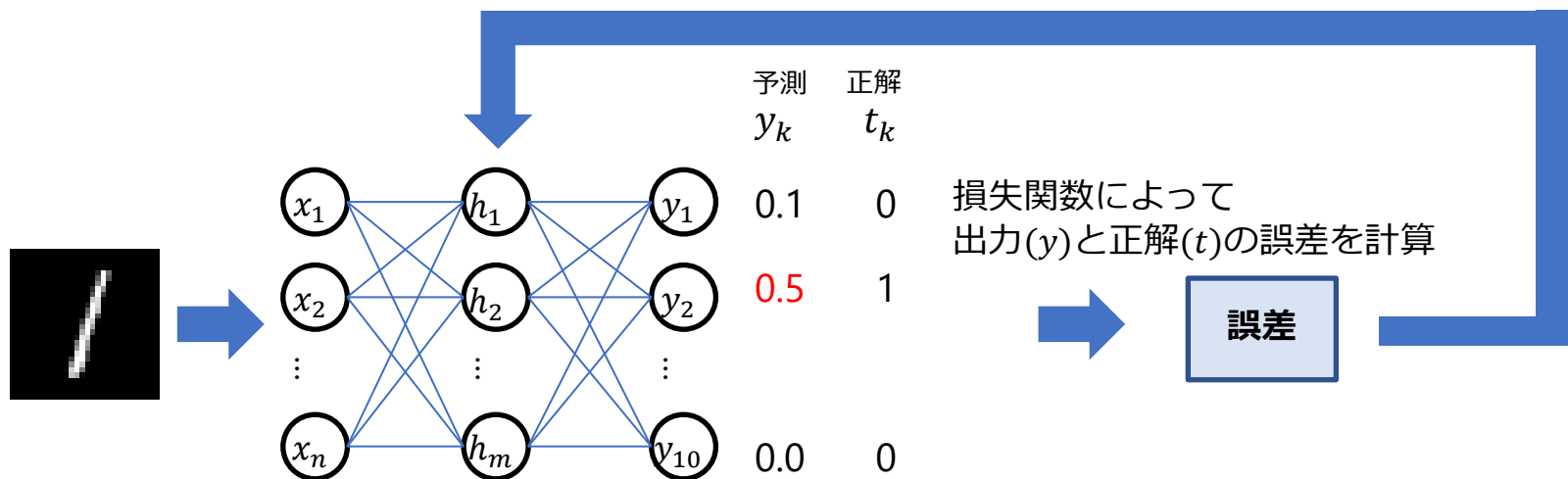
## ❖ ニューラルネットワークの学習とは

- 損失関数を最小化するパラメータ(テンソル $W$ )を求めること

## ❖ 損失関数

- 出力の値( $y$ )と正解の値( $t$ )の誤差を表す関数
  - 間違ってる場合には大きい値、正しい場合には小さい値をとる関数を設定

誤差が小さくなるようにニューラルネットワークのパラメータ(テンソル $W$ )を修正



- ❖ 学習には以下の3つのデータが必要なため、持っているデータを分割する必要がある
  - **訓練データ**：モデルのパラメータを見つけるためのデータ
    - 主に全データの6～8割が用いられる
  - **検証データ**：ハイパーパラメータ(人が決定すべき項目)を検証するためのデータ
    - 全データの1～2割程度
  - **テストデータ**：学習済みモデルの汎化性能を評価するためのデータ(実運用で想定される精度がここでわかる)
    - 全データの1～2割程度

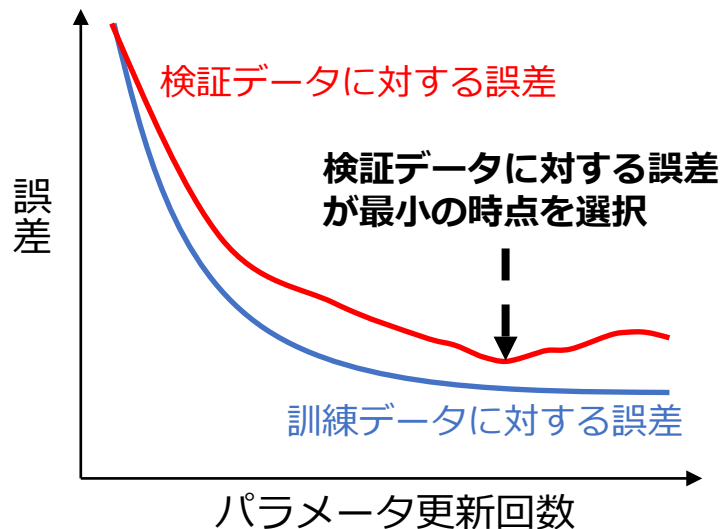


## ❖ 訓練データの一部を検証データとして使用

- 検証データはモデルの訓練自体には使用しない
- 検証データはモデルの評価(誤差、精度計算)に使用する
  - テストデータとは別

### 検証データの使用例

- ・ ニューラルネットワークの学習推移を確認
- ・ 複数のモデルを比較し採用するモデルを選択

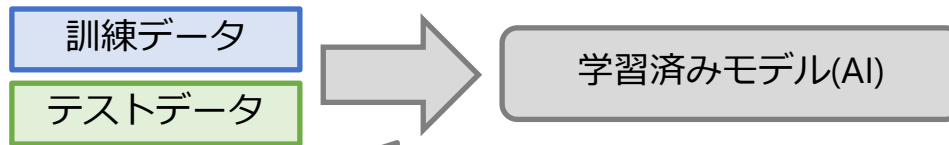


モデル	検証誤差
モデルA	0.2
モデルB	0.1
モデルC	0.3
モデルD	0.5

← 採用！

## ❖ 学習の真の目的

- 訓練データだけではなく、未知のデータに対しても正しく予測できるようにすること

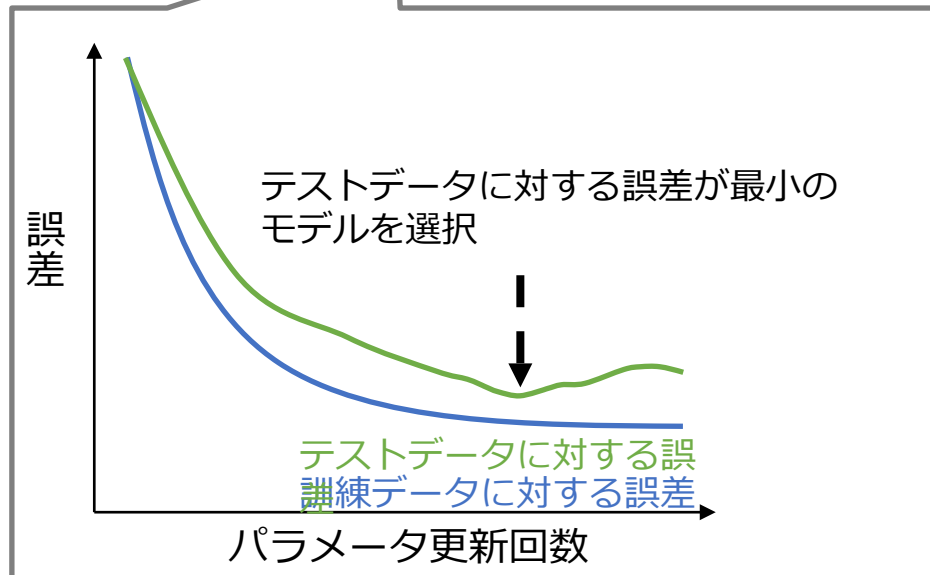


テストデータに対する精度：90%

**この学習済みモデルは未知のデータに対しても90%の精度を示すと言えるか？**

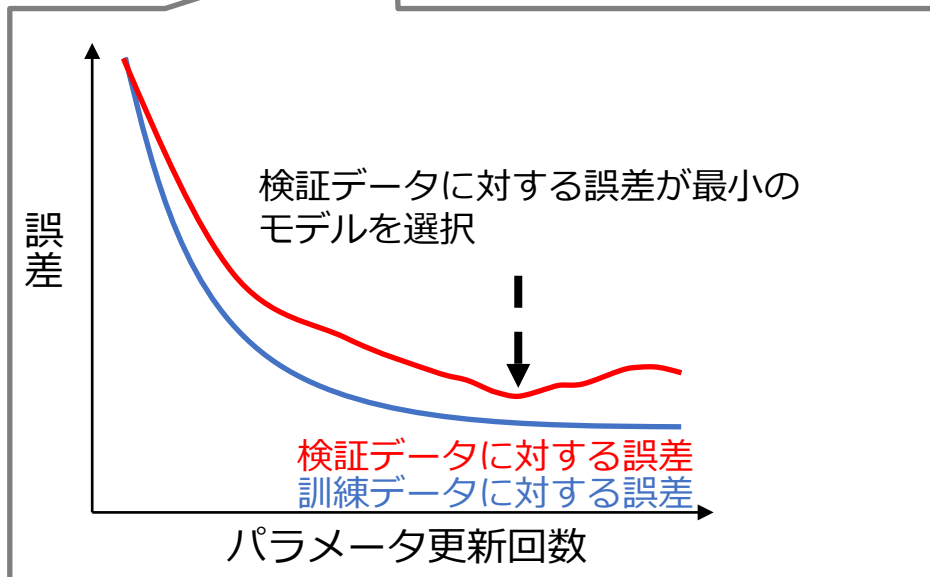
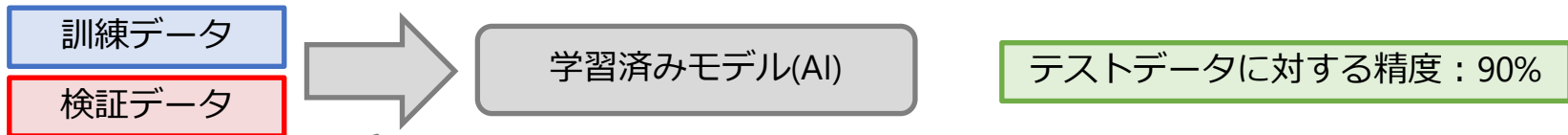
テストデータに対して良い精度を示すモデルを選択してきている

テストデータはモデルの学習、選択には使用してはいけない



## ❖ 学習の真の目的

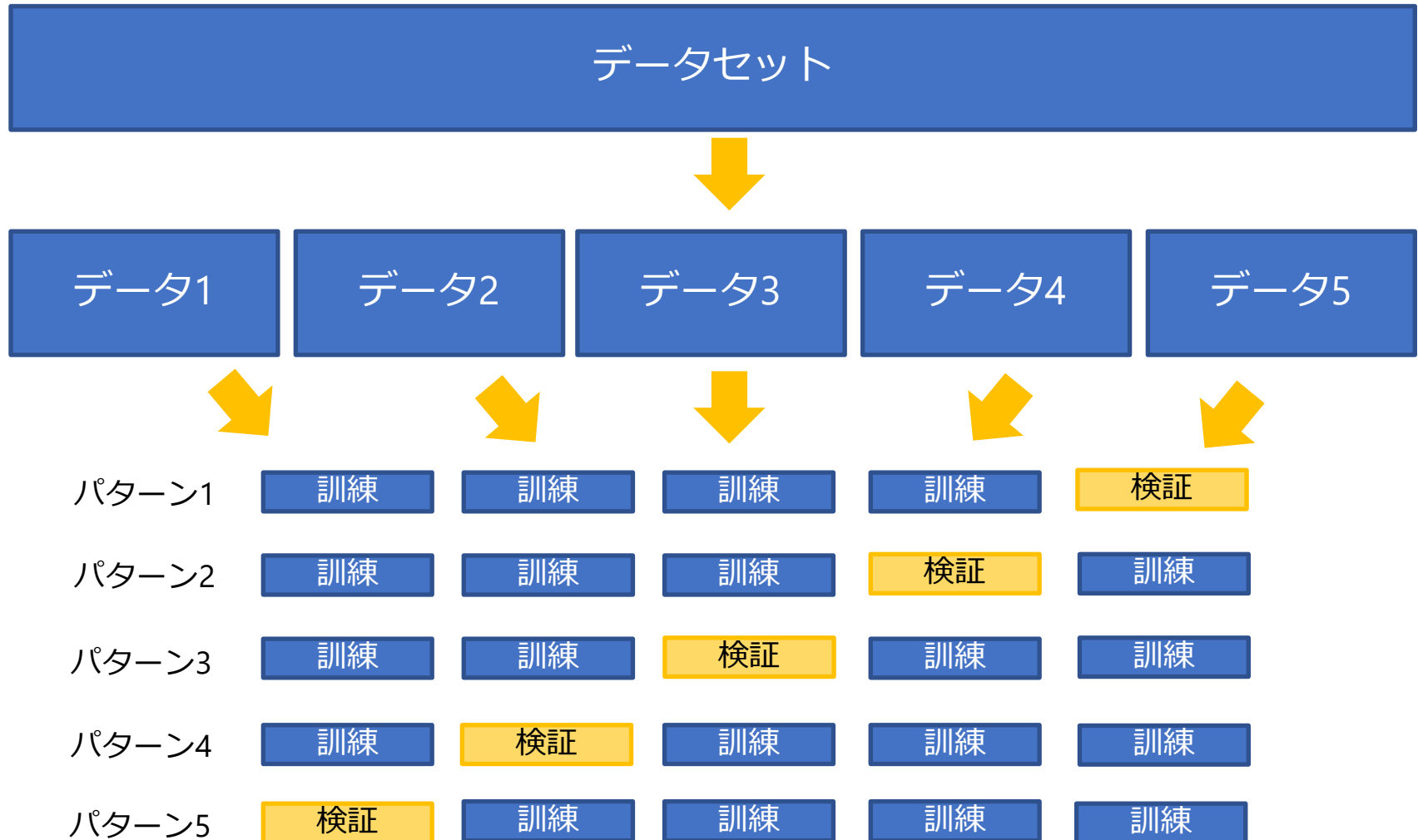
- 訓練データだけではなく、未知のデータに対しても正しく予測できるようにすること



モデルの学習に一切使用していない  
テストデータで性能を評価することで  
未知のデータに対する性能を評価する

(学習中のモデル評価やモデルの選択には  
検証データを使用する必要がある)

- ❖ 分割されたデータは1度は必ず検証する





## ❖ 分割されたデータは1度は必ず検証する

ポイント

- ❖ 1パターンのみで実験して結果が良くても、たまたまデータの組み合わせが良いだけだったかも？
- ❖ 正しい汎化性能の検証は必ずたくさんのパターンを比較すること！

データ1

パターン1	訓練	訓練	訓練	訓練	検証
パターン2	訓練	訓練	訓練	検証	訓練
パターン3	訓練	訓練	検証	訓練	訓練
パターン4	訓練	検証	訓練	訓練	訓練
パターン5	検証	訓練	訓練	訓練	訓練

# 評価指標

10

- 分類問題における機械学習モデルの推定機能を評価する指標
  - 唯一の指標では評価できない
- 評価指標
  - 正解率 : Accuracy
  - 適合率 : Precision
  - 再現率 (感度) : Recall, True Positive Rate(TPR)
  - F値 : F-measure
  - 特異度 : specificity, True Negative Rate(TNR)
  - 偽陽性率 : False Positive Rate(FPR)

# 評価指標の準備

11

- 混同行列(Confusion matrix)

		機械学習モデルの推定結果	
		陰性(負)	陽性(正)
真値	陰性(負)	真陰性 : True Negative(TN)	偽陽性 : False Positive(TN)
	陽性(正)	偽陰性 : False Negative(FN)	真陽性 : True Positive(TN)

- 実際の値(真値)と推定結果の比較
  - 上記は2クラス分類の混同行列
  - 多クラス分類の混同行列もある

# 評価指標の準備

12

- 混同行列(Confusion matrix)

		機械学習モデルの推定結果	
		陽性(正)	陰性(負)
真値	陽性(正)	真陽性 : True Positive(TP)	偽陰性 : False Negative(FN)
	陰性(負)	偽陽性 : False Positive(FP)	真陰性 : True Negative(TN)

- 実際の値(真値)と推定結果の比較
  - 上記は2クラス分類の混同行列
  - 多クラス分類の混同行列もある

# 評価指標の準備

13

## • 混同行列(Confusion matrix)

		機械学習モデルの推定結果	
		陽性(正)	陰性(負)
真値	陽性(正)	真陽性 : True Positive(TP)	偽陰性 : False Negative(FN)
	陰性(負)	偽陽性 : False Positive(FP)	真陰性 : True Negative(TN)

正解 : True

- 真陽性 : True Positive(TP)
  - 分類対象を「分類対象である」と推定した数
- 真陰性 : True Negative(TN)
  - 分類対象以外のものを「分類対象以外のもの」と推定した回数
- 偽陽性 : False Positive(FP)
  - 分類対象以外のものを「分類対象」と推定した回数
- 偽陰性 : False Negative(FN)
  - 分類対象を「分類対象以外のもの」と推定した回数

# 評価指標の準備

14

## • 混同行列(Confusion matrix)

		機械学習モデルの推定結果	
		陰性(負)	陽性(正)
真値	陰性(負)	真陽性 : True Positive(TP)	偽陽性 : False Positive(FP)
	陽性(正)	偽陽性 : False Positive(FP)	真陽性 : True Positive(TP)

- 真陽性 : True Positive(TP)
  - 分類対象を「分類対象である」と推定した数
- 真陰性 : True Negative(TN)
  - 分類対象以外のものを「分類対象以外のもの」と推定した回数
- 偽陽性 : False Positive(FP)
  - 分類対象以外のものを「分類対象」と推定した回数
- 偽陰性 : False Negative(FN)
  - 分類対象を「分類対象以外のもの」と推定した回数

不正解 : False

# 評価指標の準備

15

- 混同行列(Confusion matrix)



「画像に映っている動物を 猫とそれ以外の動物に分類」		機械学習モデルの推定結果	
		陽性(正)	陰性(負)
真値	陽性(正)	真陽性： True Positive(TP)	偽陰性： False Negative(FN)
	陰性(負)	偽陽性： False Positive(FP)	真陰性： True Negative(TN)

- 真陽性：True Positive(TP)
  - 猫の映っている画像を「猫が映っている」と推定した数
- 真陰性：True Negative(TN)
  - 猫以外の動物が映っている画像を「猫以外の動物が映っている画像」と推定した数
- 偽陽性：False Positive(FP)
  - 猫以外の動物が映っている画像を「猫が映っている画像」と推定した数
- 偽陰性：False Negative(FN)
  - 猫が映っている画像を「猫以外の動物が映っている画像」と推定した数

# 評価指標

16

- 混同行列(Confusion matrix)

		機械学習モデルの推定結果	
		陽性(正)	陰性(負)
真値	陽性(正)	真陽性 : True Positive(TP)	偽陰性 : False Negative(FN)
	陰性(負)	偽陽性 : False Positive(FP)	真陰性 : True Negative(TN)

- 正解率 : Accuracy)
- 適合率 : Precision
- 再現率 (感度) : Recall, True Positive Rate(TPR)
- F値 : F-measure
- 特異度 : specificity, True Negative Rate(TNR)
- 偽陽性率 : False Positive Rate(FPR)



# 評価指標

17

- 混同行列(Confusion matrix)

		機械学習モデルの推定結果	
		陽性(正)	陰性(負)
真値	陽性(正)	真陽性 : True Positive(TP)	偽陰性 : False Negative(FN)
	陰性(負)	偽陽性 : False Positive(FP)	真陰性 : True Negative(TN)

- 正解率 : Accuracy
  - 全ての推定結果で正解した予測の割合

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

- 正例と負例の数の偏りが大きい場合にあまり有効ではない

# 評価指標

18

- 適用事例：天気予報
  - 日本では晴れの日が70%

		機械学習モデルの推定結果	
		晴れ	晴れ以外
真値	晴れ	70	0
	晴れ以外	30	0

- 何も考えずに常に「晴れ」と出力しても正解率は70%

# 評価指標

19

- 混同行列(Confusion matrix)

		機械学習モデルの推定結果	
		陽性(正)	陰性(負)
真値	陽性(正)	真陽性 : True Positive(TP)	偽陰性 : False Negative(FN)
	陰性(負)	偽陽性 : False Positive(FP)	真陰性 : True Negative(TN)

- 適合率 : Precision
  - 陽性と推定した中で、実際に陽性である割合
  - 偽陽性(FP)を小さくすることを重視
    - 間違って陽性と判断しては困る場合に利用

$$Precision = \frac{TP}{TP + FP}$$

# 評価指標

20

- 混同行列(Confusion matrix)

		機械学習モデルの推定結果	
		陽性(正) : 良品	陰性(負) : 非良品
真値	陽性(正) : 良品	真陽性 : True Positive(TP)	偽陰性 : False Negative(FN)
	陰性(負) : 非良品	偽陽性 : False Positive(FP)	真陰性 : True Negative(TN)

- 適合率 : Precision

- 陽性と推定した中で、実際に陽性である割合
- 偽陽性(FP)を小さくすることを重視
  - 間違って陽性と判断しては困る場合に利用

$$Precision = \frac{TP}{TP + FP}$$

例) ブランド食材(良品)の選別

- 陽性判定で高く売る
  - 間違うとクレーム
    - 非良品が高く売られる
- 陰性判定で安く売る
  - 間違ってもクレームなし
    - 良品が低価格で売られる

# 評価指標

21

- 混同行列(Confusion matrix)

		機械学習モデルの推定結果	
		陽性(正)	陰性(負)
真値	陽性(正)	真陽性 : True Positive(TP)	偽陰性 : False Negative(FN)
	陰性(負)	偽陽性 : False Positive(FP)	真陰性 : True Negative(TN)

- 再現率 (感度) : Recall, True Positive Rate(TPR)
  - 実際に陽性であるものの中で、正しく陽性と推定できた割合
  - 偽陰性(FN)を小さくすることを重視
    - 間違って陰性と判断しては困る場合に利用

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

# 評価指標

22

## • 混同行列(Confusion matrix)

		機械学習モデルの推定結果	
		陽性(正)	陰性(負)
真値	陽性(正)	真陽性 : True Positive(TP)	偽陰性 : False Negative(FN)
	陰性(負)	偽陽性 : False Positive(FP)	真陰性 : True Negative(TN)

## • 再現率 (感度) : Recall, True Positive Rate(TPR)

- 実際に陽性であるものの中で、正しく陽性と推定できた割合
- 偽陰性(FN)を小さくすることを重視
  - 間違って陰性と判断しては困る場合に利用

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

病気の診断で重要

- 陰性で陽性判定  
→ もう一度検査可能
- 陽性で陰性判定  
→ 通常生活で病状悪化・感染拡大

- 混同行列(Confusion matrix)

		機械学習モデルの推定結果	
		陽性(正)	陰性(負)
真値	陽性(正)	真陽性 : True Positive(TP)	偽陰性 : False Negative(FN)
	陰性(負)	偽陽性 : False Positive(FP)	真陰性 : True Negative(TN)

- F値 : F-measure

- 適合率と再現率の調和平均

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

- 適合率と再現率の統合

- どちらかが極端に悪いと小さくなる

# 評価指標

24

- 混同行列(Confusion matrix)

		機械学習モデルの推定結果	
		陽性(正)	陰性(負)
真値	陽性(正)	真陽性 : True Positive(TP)	偽陰性 : False Negative(FN)
	陰性(負)	偽陽性 : False Positive(FP)	真陰性 : True Negative(TN)

- 特異度 : specificity, True Negative Rate(TNR)
  - 実際に陰性であるもののの中で、正しく陰性と推定した割合
  - 偽陽性(FP)を減らすことを重視(小さい方が良い)

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



# 評価指標

25

- 混同行列(Confusion matrix)

		機械学習モデルの推定結果	
		陽性(正)	陰性(負)
真値	陽性(正)	真陽性 : True Positive(TP)	偽陰性 : False Negative(FN)
	陰性(負)	偽陽性 : False Positive(FP)	真陰性 : True Negative(TN)

- 偽陽性率 : False Positive Rate(FPR)
  - 実際に陰性であるもののの中で、間違って陽性と推定した割合
    - 小さい方が良い : 1- (特異度) で計算可能

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

❖ **The CIFAR-10 dataset** : 主に画像認識を目的としたチュートリアル的なオープンデータセット。より複雑なCIFAR-100も提供されている

- 32x32ピクセルサイズ
- 6万枚
- カラー画像
- 10クラス

- ラベル「0」 : airplane (飛行機)
- ラベル「1」 : automobile (自動車)
- ラベル「2」 : bird (鳥)
- ラベル「3」 : cat (猫)
- ラベル「4」 : deer (鹿)
- ラベル「5」 : dog (犬)
- ラベル「6」 : frog (カエル)
- ラベル「7」 : horse (馬)
- ラベル「8」 : ship (船)
- ラベル「9」 : truck (トラック)

airplane



automobile



bird



cat



deer



dog



frog



horse



ship



truck



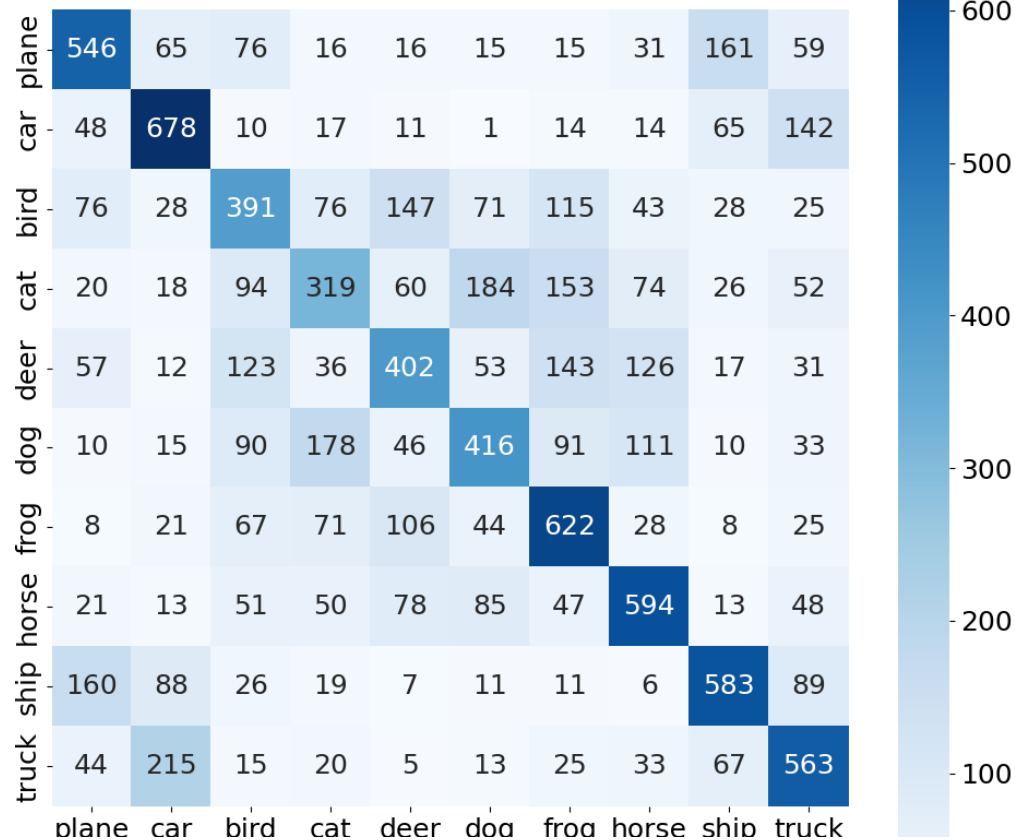
CIFAR-10 :

<https://www.cs.toronto.edu/~kriz/cifar.html>

❖ 混同行列

❖ 精度51%

正解クラス



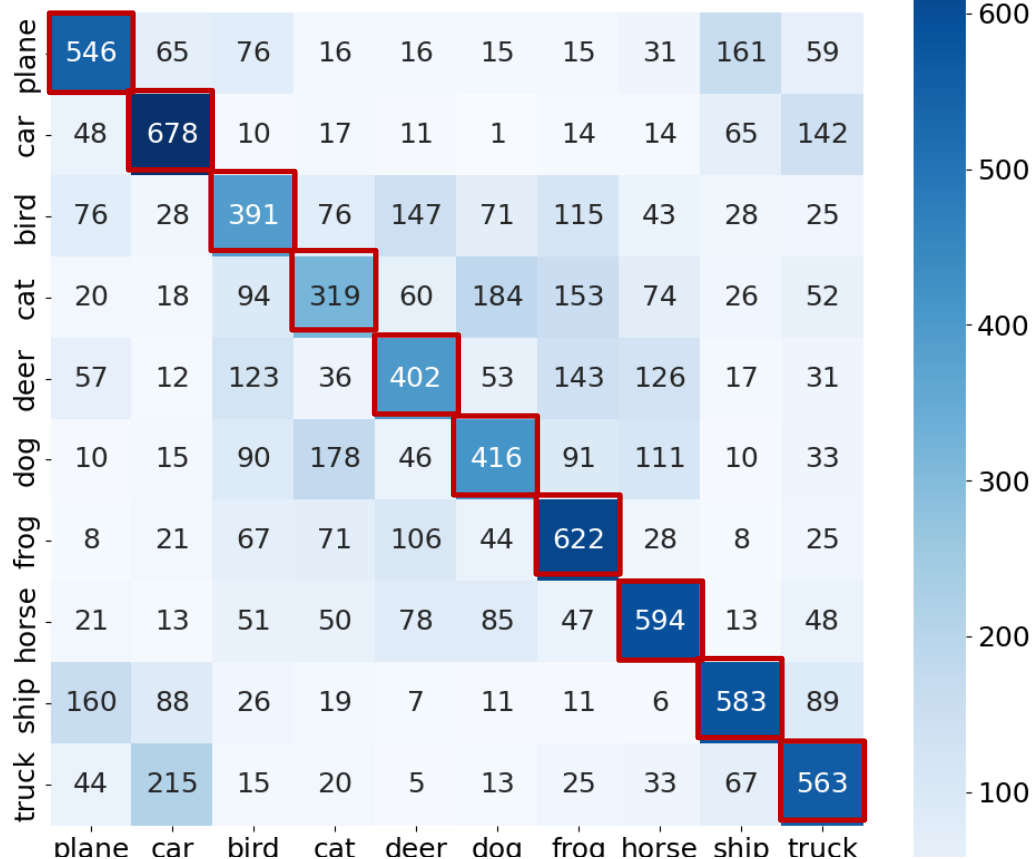
推定クラス

❖ 混同行列

対角線上は正しく分類できた  
画像枚数を表しています

❖ 精度51%

正解クラス



推定クラス

## ❖ 混同行列

❖ 精度51%

正解クラス



推定クラス