

AI ジェネラリスト基礎講座

第 2 章：機械学習

- AI 活用を進めるには、AI にやらせたいことをタスクという言葉で適切に定義できることが重要
- AI によって解決できそうな課題を考え、タスクに分解してみよう

課題 (記入してください)	タスク (記入してください)
例、自動運転車	人物検出+標識分類+30分後天候予測+1時間後渋滞予測+...

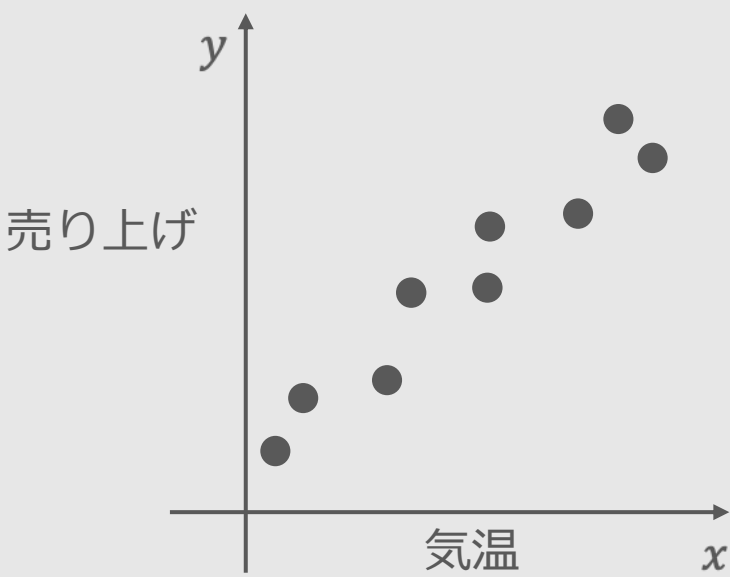
- 教師あり学習、教師なし学習、強化学習の違いを整理しよう
- それぞれについて、具体的なタスクも考えてみよう

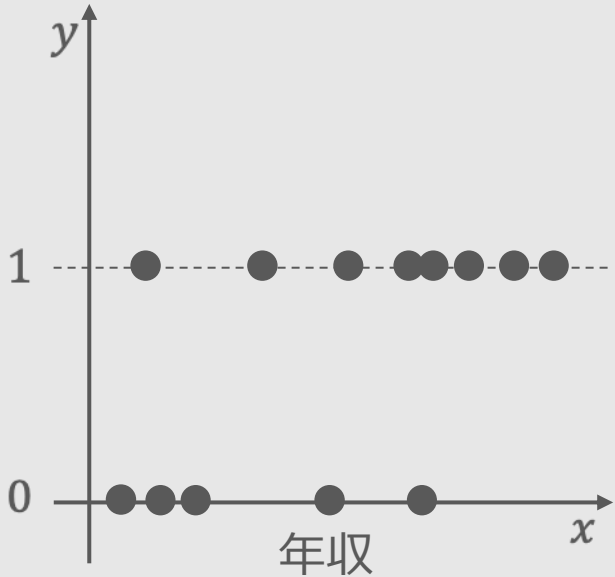
学習方法	説明 (記入してください)	具体的なタスク (記入してください)
教師あり 学習		
教師なし 学習		
強化 学習		

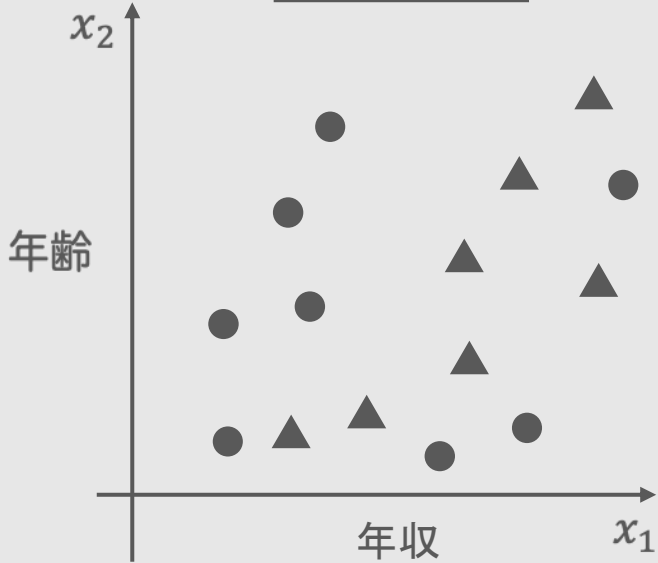
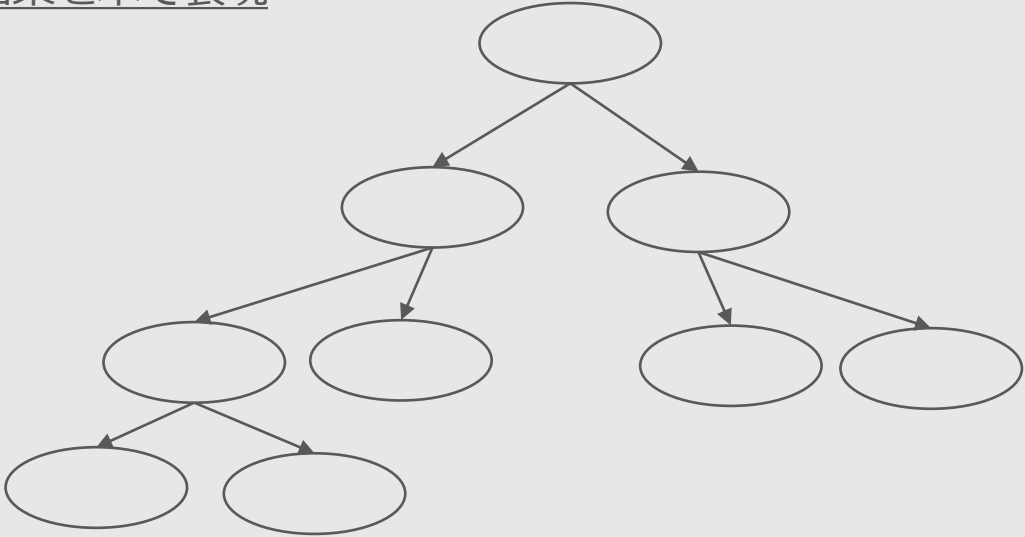
- 教師あり学習は、回帰と分類に分かれる
- 回帰と分類の違いを整理しよう

学習方法	説明(記入してください)
回帰	
分類	


- 主要な手法について、図を描くことによって理解を深めよう

手法名	線形回帰
学習方法 の分類	教師あり(回帰) or 教師あり(分類) or 教師なし
図	 <p>A scatter plot illustrating a positive linear relationship. The horizontal axis (x-axis) is labeled '気温' (Temperature) and the vertical axis (y-axis) is labeled '売り上げ' (Sales Volume). There are 10 data points plotted, showing a clear upward trend where higher temperatures correspond to higher sales volumes. The axes are labeled with 'x' and 'y' at their respective ends.</p>
特徴	
学習の 対象	

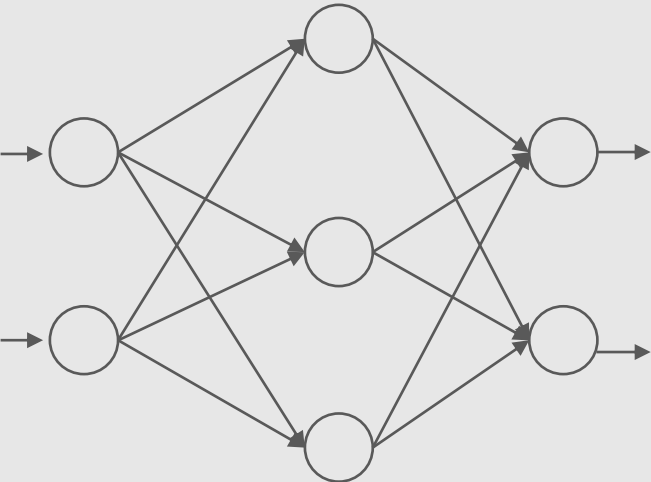
手法名	ロジスティック回帰	
学習方法 の分類	教師あり(回帰) or 教師あり(分類) or 教師なし	
図		<p>σ : _____ 関数</p> <p>w_1 : 関数σの _____ を決めるパラメータ</p> <p>w_0 : 関数σの _____ を決めるパラメータ</p>
	1 : 家を持っている、0 : 家を持っていない	
特徴		
学習の 対象		

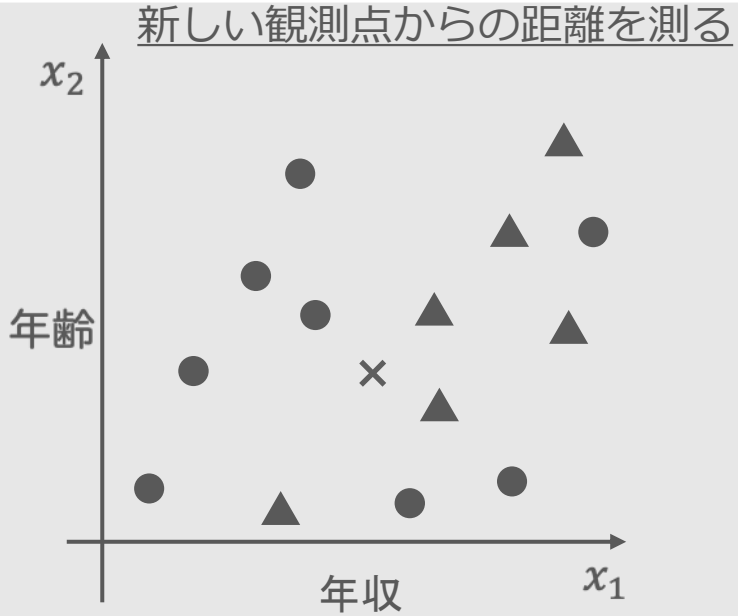
手法名	決定木
学習方法 の分類	教師あり(回帰) or 教師あり(分類) or 教師なし
図	<div><div><p>境界線を引く</p><p>● : 家を持っていない ▲ : 家を持っている</p></div><div><p>結果を木で表現</p><p>境界線は、_____が小さくなるように選ぶ。 木の深さは、_____によって決める。</p></div></div>
特徴	
学習の 対象	

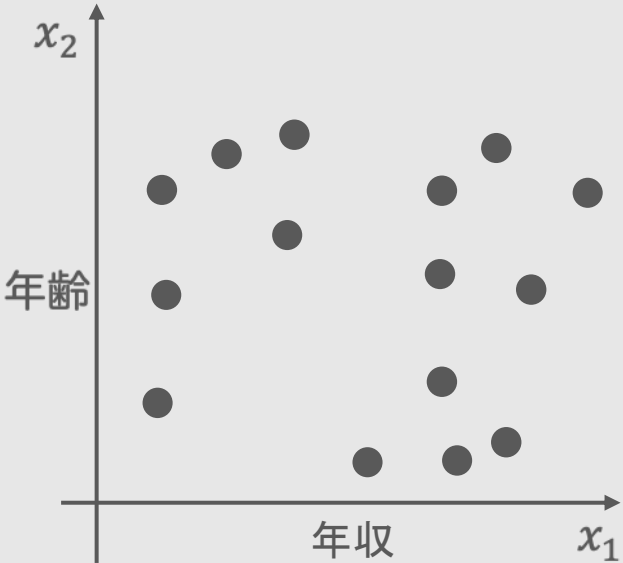
手法名	ランダムフォレスト
学習方法 の分類	教師あり(回帰) or 教師あり(分類) or 教師なし
図	<div><div><div>学習用データ</div><div><div></div><div></div><div></div><div></div></div></div><div>_____にデータを選ぶ</div><div>枝を伸ばすとき、 候補となる変数を _____に選択する</div></div> <p>木の数は、基本的に多い方がいいが、_____個くらいにすることが多い。 複数のモデルを構築し、それらモデルの出力の平均値を予測値として採用する方法のことを 一般に_____学習という</p>
特徴	
学習の 対象	

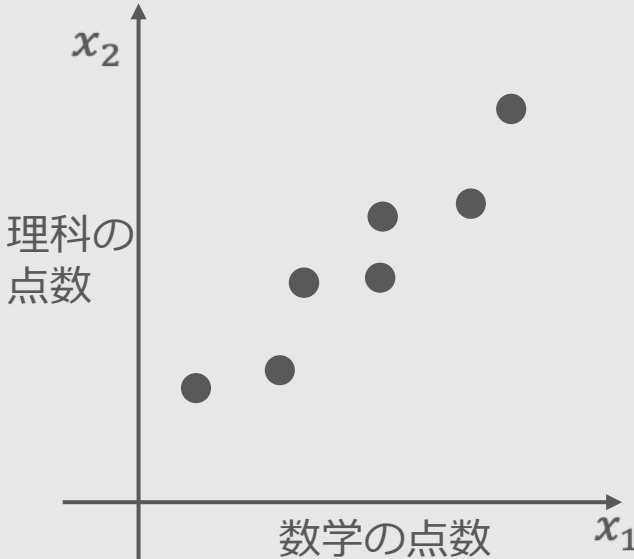
手法名	アダブースト
学習方法 の分類	教師あり(回帰) or 教師あり(分類) or 教師なし
図	<div data-bbox="746 297 2397 492"><div data-bbox="1126 325 1421 396">学習用データ</div><div data-bbox="1505 297 2397 347">木の数は、 _____ によって決める</div><div data-bbox="746 396 1905 492"></div></div>
特徴	
学習の 対象	

手法名	サポートベクトルマシン
学習方法 の分類	教師あり(回帰) or 教師あり(分類) or 教師なし
図	<div><p>境界線を引く</p><p>● : 家を持っていない ▲ : 家を持っている</p></div>
特徴	
学習の 対象	

手法名	ニューラルネットワーク（多層パーセプトロンと呼ぶこともある）
学習方法 の分類	教師あり(回帰) or 教師あり(分類) or 教師なし
図	<div>  </div> <div> $W^1 = \begin{pmatrix} w_{11}^1 & w_{12}^1 & w_{13}^1 \\ w_{21}^1 & w_{22}^1 & w_{23}^1 \end{pmatrix}, W^2 = \begin{pmatrix} w_{11}^2 & w_{12}^2 \\ w_{21}^2 & w_{22}^2 \\ w_{31}^2 & w_{32}^2 \end{pmatrix}$ </div> <div> <p>この図の場合、重み(パラメータ)の数は、合計_____個</p> </div>
特徴	
学習の 対象	

手法名	k近傍法
学習方法 の分類	教師あり(回帰) or 教師あり(分類) or 教師なし
図	<div data-bbox="509 282 1243 892"><p>新しい観測点からの距離を測る</p><p>● : 家を持っていない ▲ : 家を持っている</p></div> <div data-bbox="1472 325 2272 714"><p>$k = 5$の場合 新しい観測点xの予測結果は、____</p><p>$k = 3$の場合 新しい観測点xの予測結果は、____</p><p>kは、_____によって決める</p></div>
特徴	
学習の 対象	

手法名	k-means法
学習方法 の分類	教師あり(回帰) or 教師あり(分類) or 教師なし
図	<p data-bbox="504 301 1217 347"><u>グルーピング(クラスタリング)を行う</u></p> <div data-bbox="466 401 1085 958"><p>The scatter plot shows two clusters of data points. The horizontal axis is labeled x_1 (年収) and the vertical axis is labeled x_2 (年齢). There are 15 data points in total, distributed across the two clusters.</p></div> <p data-bbox="1360 439 2303 482">kは、 _____</p> <p data-bbox="1360 586 1681 629">によって決める。</p> <p data-bbox="1360 739 2303 782">各クラスタの中心点は、 _____</p> <p data-bbox="1360 892 1979 935">が最も小さくなるように求める。</p>
特徴	
学習の 対象	

手法名	主成分分析																																			
学習方法 の分類	教師あり(回帰) or 教師あり(分類) or 教師なし																																			
図		<table><tr><th>x_1</th><th>x_2</th><th>z_1</th><th>z_2</th></tr><tr><td>5</td><td>20</td><td></td><td></td></tr><tr><td>20</td><td>25</td><td></td><td></td></tr><tr><td>45</td><td>50</td><td></td><td></td></tr><tr><td>65</td><td>50</td><td></td><td></td></tr><tr><td>70</td><td>75</td><td></td><td></td></tr><tr><td>80</td><td>80</td><td></td><td></td></tr><tr><td>90</td><td>90</td><td></td><td></td></tr></table>	x_1	x_2	z_1	z_2	5	20			20	25			45	50			65	50			70	75			80	80			90	90			変数の数が多くなると、計算負荷が指数関数的に増える。これを_____という。主成分分析を用いると、変数の数を減らすことができる。	
	x_1	x_2	z_1	z_2																																
5	20																																			
20	25																																			
45	50																																			
65	50																																			
70	75																																			
80	80																																			
90	90																																			
	_____よりも_____の方がデータの特徴をよく説明できる重要な軸 _____だけを採用することで、次元削減になる																																			
特徴																																				
学習の 対象																																				

- csvなどのテーブルデータを扱う場合、ランダムフォレストなどのアンサンブル系手法がよく用いられるが、線形回帰やロジスティック回帰などシンプルな手法もまだまだ現役
- シンプルな手法が未だに使われているのはなぜだろうか？
- その理由を考えてみよう

シンプルな手法は、 _____ が小さい

シンプルな手法は、 _____ をしやすい

シンプルな手法は、 _____ がおきにくい

シンプルな手法は、 _____ をほとんどしなくていい

- 回帰モデルの精度を評価する際には、MSE ・ RMSE ・ MAE の3つの指標がよく用いられる
- 以下の空欄に、MSE ・ RMSE ・ MAE のいずれかを入れることによって、3つの指標の特徴を整理しよう

_____は、平方根をとることによって、単位が元のデータの単位と一致するため、MSEよりも解釈しやすい

_____は、絶対誤差の単純な平均なので、解釈しやすい

_____は、誤差を2乗するため、大きく外したデータがあると値が大きくなりやすい

- 分類用モデルの精度を評価する際には、Accuracy ・ Recall ・ Precision の3つの指標がよく用いられる
- 以下の空欄に、Accuracy ・ Recall ・ Precision のいずれかを入れることによって、3つの指標の特徴を整理しよう

_____ は、本当に陽性であるデータを陽性であると予測できた割合

_____ は、正解したデータ数を全データ数で割った値

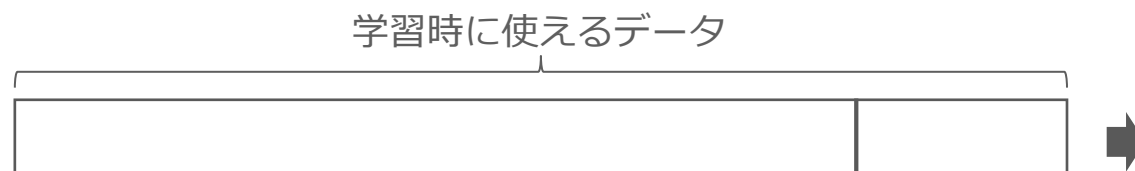
_____ は、陽性であると予測したデータのうち、本当に陽性であった割合

- モデルの性能を評価することをモデルの検証という
- モデルの検証においては、以下の7つの単語が重要
 - 訓練誤差
 - 汎化誤差
 - 過学習
 - 未学習
 - ホールドアウト法
 - 交差検証法
 - 正則化
- 次頁の空欄に7つの単語のいずれかを入れることによってそれぞれの意味を確認しよう

- 学習に用いたデータで測った誤差のことを_____という
- 学習に用いていないデータで測った誤差のことを_____という
- 学習に用いたデータに適合しすぎることを_____という
 - この場合、_____が _____よりも大きくなる
- 学習に用いたデータに適合できていないことを_____という
 - この場合、訓練誤差と汎化誤差の両方が大きくなる
- ホールドアウト法は、_____を推定するために用いられる
- _____を用いると、 _____よりも精度良く _____を推定できる
- _____を抑えるための方法として、 _____という技法がある
- 学習を行うことは訓練とも呼ばれ、汎化誤差を推定することはテストとも呼ばれる

■ ホールドアウト法と交差検証法の仕組みを確認しよう

ホールドアウト法



交差検証法(k=5の場合)



- 学習によって最適化される変数をパラメータといい、学習する前に分析者が決めておかなければならない変数をハイパーパラメータという
- 最適なハイパーパラメータを決める際に、交点探索(グリッドサーチ)という方法がよく用いられる
- 交点探索の仕組みを確認しよう

交点探索 (ニューラルネットワークを例に)



左図の場合、ハイパーパラメータの組み合わせは、
合計_____パターンある

この1つ1つの組み合わせについて、ホールドアウト法
または交差検証法を実施し、_____が
最も _____なる組み合わせを採用する

- 何らかの強化学習タスクを想定し、各基本概念に対応する具体例を記載しよう
- 想定するタスク: ()

概念	各概念に対応する具体例(記入してください)
エージェント	
行動	
環境	
方策	
状態	
報酬	
エピソード	

- 7つの基本概念の関係性を1つのフロー図で表現しよう
 - 下図の白枠に7つの基本概念のいずれかを当てはめよう

