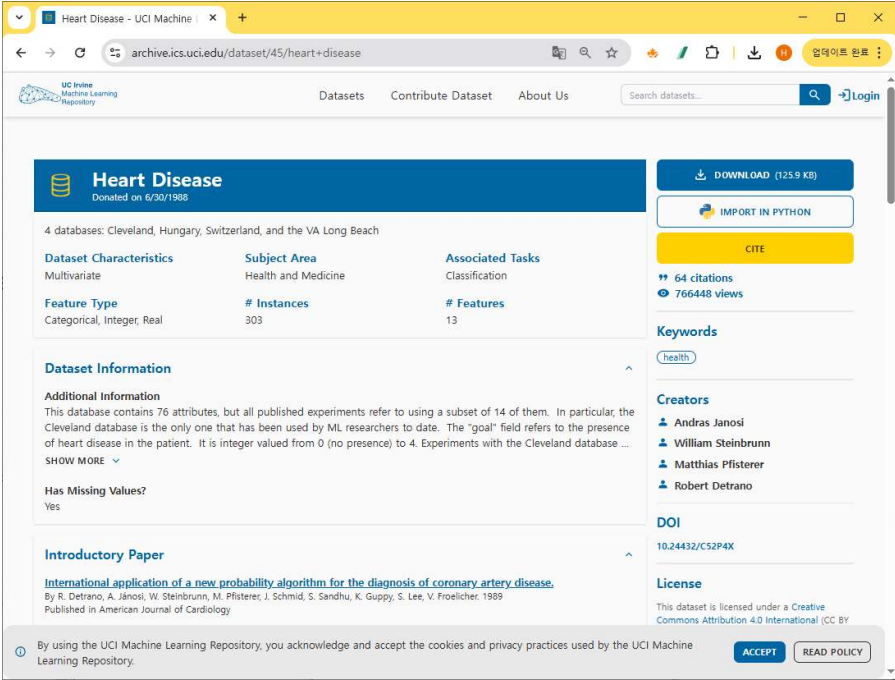


과정명		PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명
➤Intro		<ul style="list-style-type: none"><li>• 병원 기록 속에 숨어 있는 데이터, AI가 읽어내면 달라집니다</li></ul> <p>오늘날 우리는 병원에서 측정된 수많은 수치와 문진 결과들이 '차트'라는 이름으로 남겨진다는 사실에 익숙합니다. 하지만 이 데이터들은 단순히 보관되기 위한 것이 아닙니다. 예를 들어, 60세 남성 환자의 혈압, 심전도, 심박수, 협심증 여부와 같은 정보를 조합해 심장병 유무를 예측할 수 있다면 어떨까요? 바로 이런 가능성을 실현시키는 도구가 머신러닝이며, 그 핵심에 데이터 전처리와 정규화가 있습니다. AI는 마법이 아닙니다. 제대로 준비된 데이터에서 출발할 때에만 그 성능을 온전히 발휘합니다. 이번 시간에는 심장병 데이터를 바탕으로 어떻게 데이터를 가공하고, AI가 읽기 좋은 형태로 만드는지 직접 경험해보실 수 있습니다.</p>			① 본 학습 내용으로 들어가기 전, 학습 주제의 흥미를 이끌 만한 도입부의 내용이 있다면 제시해주세요.
•학습열기					② ex. 관련 뉴스기사, 실생활과 관련된 이야기 등
•학습목표					③ 저작권 침해가 되지 않도록 내용을 구성해 주세요.
➤ 학습하기					④ 출처가 있을 경우 반드시 작성해 주세요.
1. 코딩 환경 구축					
2. 심장병 데이터 전처리					
3. 데이터 정규화 및 시각화					
➤적용하기		용어설명			
➤ Outro					
•문제풀기					
내레이션					3

과정명		PyTorch로 배우는 머신러닝 알고리즘		회차명	4	화면설명
▶Intro •학습열기 •학습목표  ▶학습하기 1. 코딩 환경 구축 2. 심장병 데이터 전처리 3. 데이터 정규화 및 시각화      ▶적용하기  ▶Outro •문제풀기	<div>◆ 학습목표</div> <div>1. 실습을 위한 환경을 설정하고 필요한 패키지를 가져올 수 있다.</div> <div>2. 결측값 처리 및 특성 공학을 적용할 수 있다.</div> <div>3. 데이터 정규화 기법을 비교하고 특성 분포를 시각화할 수 있다.</div> <div>◆ 학습내용</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div>					① 학습내용과 학습목표는 강의계획서와 일치해야 하며, 필요시 강의계획서를 수정할 수 있습니다.
						② 학습목표
						③ 학습내용
<div>✓ 각 레슨에 맞는 학습 목표를 2~3개 작성해 주세요.</div> <div>✓ 1회차 당 25분 분량이 되도록 2~3개 레슨으로 구성해주세요.</div> <div>✓ 학습내용과 레슨명은 일치해야 합니다.</div>						
용어설명						
내레이션	4					

과정명		PyTorch로 배우는 머신러닝 알고리즘		회차명	4	화면설명
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> <div>➤적용하기</div> <div>➤Outro</div> <div>•문제풀기</div>	<div><div>간지</div><div>코딩 환경 구축</div></div>					
	내레이션		5			

과정명	PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명
<div>➤Intro</div> <ul style="list-style-type: none"> <li>•학습열기</li> <li>•학습목표</li> </ul> <div>➤학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> <div>➤적용하기</div> <div>➤Outro</div> <ul style="list-style-type: none"> <li>•문제풀기</li> </ul>	<ul style="list-style-type: none"> <li>• 오리지널 심장병 데이터: UCI Machine Learning Repository  <a href="https://archive.ics.uci.edu/dataset/45/heart+disease">https://archive.ics.uci.edu/dataset/45/heart+disease</a></li> </ul> <p>데이터는 4개 병원에서 수집됨</p> <p>Cleveland (303개 샘플, 가장 자주 사용됨), 우리도 이 데이터 사용</p> <p>Hungarian</p> <p>Switzerland</p> <p>Long Beach VA</p> 			<div>용어설명</div>
내레이션	<div>6</div>			

과정명	PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명																																																												
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> <div>➤적용하기</div> <div>➤Outro</div> <div>•문제풀기</div>	<div>• 주요 속성</div> <table><tr><th>열 이름</th><th>설명</th><th>결측값 유무</th><th>자료형 구분</th></tr><tr><td>age</td><td>나이 (세)</td><td>없음</td><td>연속형</td></tr><tr><td>sex</td><td>성별 (1: 남성, 0: 여성)</td><td>없음</td><td>범주형</td></tr><tr><td>cp</td><td>가슴 통증 유형 (1: 무심장통 ~ 4: 심각 통증)</td><td>없음</td><td>범주형</td></tr><tr><td>trestbps</td><td>안정시 혈압 (mm Hg)</td><td>없음</td><td>연속형</td></tr><tr><td>chol</td><td>혈중 콜레스테롤 (mg/dl)</td><td>없음</td><td>연속형</td></tr><tr><td>fbs</td><td>공복 혈당 &gt; 120mg/dl 여부 (1: 참, 0: 거짓)</td><td>없음</td><td>범주형</td></tr><tr><td>restecg</td><td>안정시 심전도 결과 (0: 정상, 1~2: 이상 소견 있음)</td><td>없음</td><td>범주형</td></tr><tr><td>thalach</td><td>최대 심박수</td><td>없음</td><td>연속형</td></tr><tr><td>exang</td><td>운동 유발 협심증 (1: 있음, 0: 없음)</td><td>없음</td><td>범주형</td></tr><tr><td>oldpeak</td><td>운동에 의한 ST depression</td><td>없음</td><td>연속형</td></tr><tr><td>slope</td><td>ST segment 기울기 (0: 하강, 1: 평평, 2: 상승)</td><td>없음</td><td>범주형</td></tr><tr><td>ca</td><td>혈관 수 (0~3, 일부는 ?로 결측)</td><td>있음</td><td>연속형(정수)</td></tr><tr><td>thal</td><td>탈라숨 검사 결과 (3: 정상, 6: 결손, 7: 이상결손)</td><td>있음</td><td>범주형</td></tr><tr><td>target</td><td>심장병 유무 (0: 없음, 1~4: 있음 → 보통 이진 변환)</td><td>없음</td><td>범주형</td></tr></table>			열 이름	설명	결측값 유무	자료형 구분	age	나이 (세)	없음	연속형	sex	성별 (1: 남성, 0: 여성)	없음	범주형	cp	가슴 통증 유형 (1: 무심장통 ~ 4: 심각 통증)	없음	범주형	trestbps	안정시 혈압 (mm Hg)	없음	연속형	chol	혈중 콜레스테롤 (mg/dl)	없음	연속형	fbs	공복 혈당 > 120mg/dl 여부 (1: 참, 0: 거짓)	없음	범주형	restecg	안정시 심전도 결과 (0: 정상, 1~2: 이상 소견 있음)	없음	범주형	thalach	최대 심박수	없음	연속형	exang	운동 유발 협심증 (1: 있음, 0: 없음)	없음	범주형	oldpeak	운동에 의한 ST depression	없음	연속형	slope	ST segment 기울기 (0: 하강, 1: 평평, 2: 상승)	없음	범주형	ca	혈관 수 (0~3, 일부는 ?로 결측)	있음	연속형(정수)	thal	탈라숨 검사 결과 (3: 정상, 6: 결손, 7: 이상결손)	있음	범주형	target	심장병 유무 (0: 없음, 1~4: 있음 → 보통 이진 변환)	없음	범주형	
	열 이름	설명	결측값 유무	자료형 구분																																																												
	age	나이 (세)	없음	연속형																																																												
	sex	성별 (1: 남성, 0: 여성)	없음	범주형																																																												
	cp	가슴 통증 유형 (1: 무심장통 ~ 4: 심각 통증)	없음	범주형																																																												
	trestbps	안정시 혈압 (mm Hg)	없음	연속형																																																												
	chol	혈중 콜레스테롤 (mg/dl)	없음	연속형																																																												
	fbs	공복 혈당 > 120mg/dl 여부 (1: 참, 0: 거짓)	없음	범주형																																																												
	restecg	안정시 심전도 결과 (0: 정상, 1~2: 이상 소견 있음)	없음	범주형																																																												
	thalach	최대 심박수	없음	연속형																																																												
	exang	운동 유발 협심증 (1: 있음, 0: 없음)	없음	범주형																																																												
	oldpeak	운동에 의한 ST depression	없음	연속형																																																												
	slope	ST segment 기울기 (0: 하강, 1: 평평, 2: 상승)	없음	범주형																																																												
	ca	혈관 수 (0~3, 일부는 ?로 결측)	있음	연속형(정수)																																																												
	thal	탈라숨 검사 결과 (3: 정상, 6: 결손, 7: 이상결손)	있음	범주형																																																												
target	심장병 유무 (0: 없음, 1~4: 있음 → 보통 이진 변환)	없음	범주형																																																													
				용어설명																																																												

내레이션

7

과정명		PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명
<div> <div>➤Intro</div> <div> <div>•학습열기</div> <div>•학습목표</div> </div> <div>➤학습하기</div> <div> <div>1.코딩 환경 구축</div> <div>2.심장병 데이터 전처리</div> <div>3.데이터 정규화 및 시각화</div> </div> <div>➤적용하기</div> <div>➤Outro</div> <div>•문제풀기</div> </div>	<div> <div>• 데이터시각화 및 정규화를 위한 라이브러리 가져오기</div> <div> <div># 라이브러리 불러오기</div> <div> import pandas as pd import numpy as np import matplotlib.pyplot as plt import seaborn as sns </div> </div> <div> <div># 최소-최대 정규화, 표준 정규화 함수 사용</div> <div> from sklearn.preprocessing import MinMaxScaler, StandardScaler </div> </div> </div>				
					용어설명
내레이션					8

과정명	PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤학습하기</div> <div>1.코딩 환경 구축</div> <div>2.심장병 데이터 전처리</div> <div>3.데이터 정규화 및 시각화</div> <div>➤적용하기</div> <div>➤Outro</div> <div>•문제풀기</div>	<div>• UCI 심장병 데이터 가져오기</div> <div>UCI 데이터 저장소 패키지 설치</div> <div>UCI 심장병 데이터 직접 가져오기</div> <div># UCI 데이터 저장소 패키지 설치</div> <div>!pip install ucimlrepo</div> <div># UCI 데이터 저장소에서 심장병 데이터셋 가져오기</div> <div>from ucimlrepo import fetch_ucirepo</div> <div># fetch dataset</div> <div>heart_disease = fetch_ucirepo(id=45)</div> <div># data (as pandas dataframes)</div> <div>X = heart_disease.data.features # 13개의 특징의 데이터프레임</div> <div>y = heart_disease.data.targets # 심장병 유무(target)</div>			
				용어설명

내레이션

9

과정명	PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> <div>➤적용하기</div> <div>➤Outro</div> <div>•문제풀기</div>	<div>• 심장병 데이터프레임 X, y</div> <div><div><div>[3] # 13개의 특징의 데이터프레임 X</div><div><div><div>↻</div><div>age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal</div><div><div><div>0 63 1 1 145 233 1 2 150 0 2.3 3 0.0 6.0</div><div>1 67 1 4 160 286 0 2 108 1 1.5 2 3.0 3.0</div><div>2 67 1 4 120 229 0 2 129 1 2.6 2 2.0 7.0</div><div>3 37 1 3 130 250 0 0 187 0 3.5 3 0.0 3.0</div><div>4 41 0 2 130 204 0 2 172 0 1.4 1 0.0 3.0</div><div>... ... ... ... ... ... ... ...</div><div>298 45 1 1 110 264 0 0 132 0 1.2 2 0.0 7.0</div><div>299 68 1 4 144 193 1 0 141 0 3.4 2 2.0 7.0</div><div>300 57 1 4 130 131 0 0 115 1 1.2 2 1.0 7.0</div><div>301 57 0 2 130 236 0 2 174 0 0.0 2 1.0 3.0</div><div>302 38 1 3 138 175 0 0 173 0 0.0 1 NaN 3.0</div></div><div>303 rows × 13 columns</div></div></div></div><div><div><div>[7] # 심장병 유무(target) y</div><div><div><div>↻</div><div>num</div><div><div><div>0 0</div><div>1 2</div><div>2 1</div><div>3 0</div><div>4 0</div><div>... ...</div><div>298 1</div><div>299 2</div><div>300 3</div><div>301 1</div><div>302 0</div></div><div>303 rows × 1 columns</div></div></div></div></div></div></div></div>			
				용어설명



과정명		PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명
<div> <div>➤Intro</div> <div> <div>•학습열기</div> <div>•학습목표</div> </div> <div>➤ 학습하기</div> <div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> </div> <div>➤ 적용하기</div> <div>➤ Outro</div> <div>•문제풀기</div> </div>	<div> <div>간지</div> <div>심장병 데이터 전처리</div> </div>				
					용어설명
내레이션					11

과정명	PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명																																																																																																																																																																																				
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> <div>➤적용하기</div> <div>➤Outro</div> <div>•문제풀기</div>	<div>• 전처리를 수행할 df 생성</div> <div># X, y를 합쳐서 df으로 생성, target열을 심장병 유무(1, 0)으로 생성</div> <div>df = X.copy()</div> <div># 열 target 생성: 0이면 정상, 나머지는 심장병 있음</div> <div>df['target'] = np.where(y.num &lt; 1, 0, 1)</div> <div>df</div> <table><thead><tr><th></th><th>age</th><th>sex</th><th>cp</th><th>trestbps</th><th>chol</th><th>fbs</th><th>restecg</th><th>thalach</th><th>exang</th><th>oldpeak</th><th>slope</th><th>ca</th><th>thal</th><th>target</th></tr></thead><tbody><tr><td>0</td><td>63</td><td>1</td><td>1</td><td>145</td><td>233</td><td>1</td><td>2</td><td>150</td><td>0</td><td>2.3</td><td>3</td><td>0.0</td><td>6.0</td><td>0</td></tr><tr><td>1</td><td>67</td><td>1</td><td>4</td><td>160</td><td>286</td><td>0</td><td>2</td><td>108</td><td>1</td><td>1.5</td><td>2</td><td>3.0</td><td>3.0</td><td>1</td></tr><tr><td>2</td><td>67</td><td>1</td><td>4</td><td>120</td><td>229</td><td>0</td><td>2</td><td>129</td><td>1</td><td>2.6</td><td>2</td><td>2.0</td><td>7.0</td><td>1</td></tr><tr><td>3</td><td>37</td><td>1</td><td>3</td><td>130</td><td>250</td><td>0</td><td>0</td><td>187</td><td>0</td><td>3.5</td><td>3</td><td>0.0</td><td>3.0</td><td>0</td></tr><tr><td>4</td><td>41</td><td>0</td><td>2</td><td>130</td><td>204</td><td>0</td><td>2</td><td>172</td><td>0</td><td>1.4</td><td>1</td><td>0.0</td><td>3.0</td><td>0</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>298</td><td>45</td><td>1</td><td>1</td><td>110</td><td>264</td><td>0</td><td>0</td><td>132</td><td>0</td><td>1.2</td><td>2</td><td>0.0</td><td>7.0</td><td>1</td></tr><tr><td>299</td><td>68</td><td>1</td><td>4</td><td>144</td><td>193</td><td>1</td><td>0</td><td>141</td><td>0</td><td>3.4</td><td>2</td><td>2.0</td><td>7.0</td><td>1</td></tr><tr><td>300</td><td>57</td><td>1</td><td>4</td><td>130</td><td>131</td><td>0</td><td>0</td><td>115</td><td>1</td><td>1.2</td><td>2</td><td>1.0</td><td>7.0</td><td>1</td></tr><tr><td>301</td><td>57</td><td>0</td><td>2</td><td>130</td><td>236</td><td>0</td><td>2</td><td>174</td><td>0</td><td>0.0</td><td>2</td><td>1.0</td><td>3.0</td><td>1</td></tr><tr><td>302</td><td>38</td><td>1</td><td>3</td><td>138</td><td>175</td><td>0</td><td>0</td><td>173</td><td>0</td><td>0.0</td><td>1</td><td>NaN</td><td>3.0</td><td>0</td></tr></tbody></table> <div>303 rows × 14 columns</div>				age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target	0	63	1	1	145	233	1	2	150	0	2.3	3	0.0	6.0	0	1	67	1	4	160	286	0	2	108	1	1.5	2	3.0	3.0	1	2	67	1	4	120	229	0	2	129	1	2.6	2	2.0	7.0	1	3	37	1	3	130	250	0	0	187	0	3.5	3	0.0	3.0	0	4	41	0	2	130	204	0	2	172	0	1.4	1	0.0	3.0	0	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	298	45	1	1	110	264	0	0	132	0	1.2	2	0.0	7.0	1	299	68	1	4	144	193	1	0	141	0	3.4	2	2.0	7.0	1	300	57	1	4	130	131	0	0	115	1	1.2	2	1.0	7.0	1	301	57	0	2	130	236	0	2	174	0	0.0	2	1.0	3.0	1	302	38	1	3	138	175	0	0	173	0	0.0	1	NaN	3.0	0	
		age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target																																																																																																																																																																									
	0	63	1	1	145	233	1	2	150	0	2.3	3	0.0	6.0	0																																																																																																																																																																									
	1	67	1	4	160	286	0	2	108	1	1.5	2	3.0	3.0	1																																																																																																																																																																									
	2	67	1	4	120	229	0	2	129	1	2.6	2	2.0	7.0	1																																																																																																																																																																									
	3	37	1	3	130	250	0	0	187	0	3.5	3	0.0	3.0	0																																																																																																																																																																									
	4	41	0	2	130	204	0	2	172	0	1.4	1	0.0	3.0	0																																																																																																																																																																									
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...																																																																																																																																																																									
	298	45	1	1	110	264	0	0	132	0	1.2	2	0.0	7.0	1																																																																																																																																																																									
	299	68	1	4	144	193	1	0	141	0	3.4	2	2.0	7.0	1																																																																																																																																																																									
300	57	1	4	130	131	0	0	115	1	1.2	2	1.0	7.0	1																																																																																																																																																																										
301	57	0	2	130	236	0	2	174	0	0.0	2	1.0	3.0	1																																																																																																																																																																										
302	38	1	3	138	175	0	0	173	0	0.0	1	NaN	3.0	0																																																																																																																																																																										
				용어설명																																																																																																																																																																																				

과정명	PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명																																																												
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> <div>➤적용하기</div> <div>➤Outro</div> <div>•문제풀기</div>	<div>• 결측값(missing value) 확인</div> <div>df.info()</div> <div># 열별로 결측값 수 확인</div> <div>df.isna().sum()</div> <div># 결측값이 있는 열 이름 출력</div> <div>df.columns[df.isnull().any(axis=0)]</div> <div>Index(['ca', 'thal'], dtype='object')</div>	<div>출력</div> <div><div>age0</div><div>sex0</div><div>cp0</div><div>trestbps0</div><div>chol0</div><div>fbs0</div><div>restecg0</div><div>thalach0</div><div>exang0</div><div>oldpeak0</div><div>slope0</div><div>ca4</div><div>thal2</div><div>target0</div><div>dtype: int64</div></div>	<div>&lt;class 'pandas.core.frame.DataFrame'&gt;</div> <div>RangeIndex: 303 entries, 0 to 302</div> <div>Data columns (total 14 columns):</div> <div><table><thead><tr><th>#</th><th>Column</th><th>Non-Null Count</th><th>Dtype</th></tr></thead><tbody><tr><td>0</td><td>age</td><td>303 non-null</td><td>int64</td></tr><tr><td>1</td><td>sex</td><td>303 non-null</td><td>int64</td></tr><tr><td>2</td><td>cp</td><td>303 non-null</td><td>int64</td></tr><tr><td>3</td><td>trestbps</td><td>303 non-null</td><td>int64</td></tr><tr><td>4</td><td>chol</td><td>303 non-null</td><td>int64</td></tr><tr><td>5</td><td>fbs</td><td>303 non-null</td><td>int64</td></tr><tr><td>6</td><td>restecg</td><td>303 non-null</td><td>int64</td></tr><tr><td>7</td><td>thalach</td><td>303 non-null</td><td>int64</td></tr><tr><td>8</td><td>exang</td><td>303 non-null</td><td>int64</td></tr><tr><td>9</td><td>oldpeak</td><td>303 non-null</td><td>float64</td></tr><tr><td>10</td><td>slope</td><td>303 non-null</td><td>int64</td></tr><tr><td>11</td><td>ca</td><td>299 non-null</td><td>float64</td></tr><tr><td>12</td><td>thal</td><td>301 non-null</td><td>float64</td></tr><tr><td>13</td><td>target</td><td>303 non-null</td><td>int64</td></tr></tbody></table></div> <div>dtypes: float64(3), int64(11)</div> <div>memory usage: 33.3 KB</div>	#	Column	Non-Null Count	Dtype	0	age	303 non-null	int64	1	sex	303 non-null	int64	2	cp	303 non-null	int64	3	trestbps	303 non-null	int64	4	chol	303 non-null	int64	5	fbs	303 non-null	int64	6	restecg	303 non-null	int64	7	thalach	303 non-null	int64	8	exang	303 non-null	int64	9	oldpeak	303 non-null	float64	10	slope	303 non-null	int64	11	ca	299 non-null	float64	12	thal	301 non-null	float64	13	target	303 non-null	int64	<div>용어설명</div>
	#	Column	Non-Null Count	Dtype																																																												
0	age	303 non-null	int64																																																													
1	sex	303 non-null	int64																																																													
2	cp	303 non-null	int64																																																													
3	trestbps	303 non-null	int64																																																													
4	chol	303 non-null	int64																																																													
5	fbs	303 non-null	int64																																																													
6	restecg	303 non-null	int64																																																													
7	thalach	303 non-null	int64																																																													
8	exang	303 non-null	int64																																																													
9	oldpeak	303 non-null	float64																																																													
10	slope	303 non-null	int64																																																													
11	ca	299 non-null	float64																																																													
12	thal	301 non-null	float64																																																													
13	target	303 non-null	int64																																																													

과정명	PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명																																																																																																																					
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤ 학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> <div>➤ 적용하기</div> <div>➤ Outro</div> <div>•문제풀기</div>	<div>• 하나라도 결측값이 있는 행 확인</div> <table><thead><tr><th>열 이름</th><th>설명</th><th>결측값 유무</th><th>자료형 구분</th></tr></thead><tbody><tr><td>ca</td><td>혈관 수 (0~3, 일부는 ?로 결측)</td><td>있음</td><td>연속형(정수)</td></tr><tr><td>thal</td><td>탈라숨 검사 결과 (3: 정상, 6: 결손, 7: 이상결손)</td><td>있음</td><td>범주형</td></tr></tbody></table> <div># 하나라도 결측값이 있는 행 확인</div> <div>df[df.isna().any(axis=1)]</div> <div>[13] # 하나라도 결측값이 있는 행 확인</div> <div>df[df.isna().any(axis=1)]</div> <div><div>↺</div><table><thead><tr><th></th><th>age</th><th>sex</th><th>cp</th><th>trestbps</th><th>chol</th><th>fbs</th><th>restecg</th><th>thalach</th><th>exang</th><th>oldpeak</th><th>slope</th><th>ca</th><th>thal</th><th>target</th></tr></thead><tbody><tr><td>87</td><td>53</td><td>0</td><td>3</td><td>128</td><td>216</td><td>0</td><td>2</td><td>115</td><td>0</td><td>0.0</td><td>1</td><td>0.0</td><td>NaN</td><td>0</td></tr><tr><td>166</td><td>52</td><td>1</td><td>3</td><td>138</td><td>223</td><td>0</td><td>0</td><td>169</td><td>0</td><td>0.0</td><td>1</td><td>NaN</td><td>3.0</td><td>0</td></tr><tr><td>192</td><td>43</td><td>1</td><td>4</td><td>132</td><td>247</td><td>1</td><td>2</td><td>143</td><td>1</td><td>0.1</td><td>2</td><td>NaN</td><td>7.0</td><td>1</td></tr><tr><td>266</td><td>52</td><td>1</td><td>4</td><td>128</td><td>204</td><td>1</td><td>0</td><td>156</td><td>1</td><td>1.0</td><td>2</td><td>0.0</td><td>NaN</td><td>1</td></tr><tr><td>287</td><td>58</td><td>1</td><td>2</td><td>125</td><td>220</td><td>0</td><td>0</td><td>144</td><td>0</td><td>0.4</td><td>2</td><td>NaN</td><td>7.0</td><td>0</td></tr><tr><td>302</td><td>38</td><td>1</td><td>3</td><td>138</td><td>175</td><td>0</td><td>0</td><td>173</td><td>0</td><td>0.0</td><td>1</td><td>NaN</td><td>3.0</td><td>0</td></tr></tbody></table><div><div>📊</div><div>📈</div></div></div>			열 이름	설명	결측값 유무	자료형 구분	ca	혈관 수 (0~3, 일부는 ?로 결측)	있음	연속형(정수)	thal	탈라숨 검사 결과 (3: 정상, 6: 결손, 7: 이상결손)	있음	범주형		age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target	87	53	0	3	128	216	0	2	115	0	0.0	1	0.0	NaN	0	166	52	1	3	138	223	0	0	169	0	0.0	1	NaN	3.0	0	192	43	1	4	132	247	1	2	143	1	0.1	2	NaN	7.0	1	266	52	1	4	128	204	1	0	156	1	1.0	2	0.0	NaN	1	287	58	1	2	125	220	0	0	144	0	0.4	2	NaN	7.0	0	302	38	1	3	138	175	0	0	173	0	0.0	1	NaN	3.0	0	
	열 이름	설명	결측값 유무	자료형 구분																																																																																																																					
	ca	혈관 수 (0~3, 일부는 ?로 결측)	있음	연속형(정수)																																																																																																																					
	thal	탈라숨 검사 결과 (3: 정상, 6: 결손, 7: 이상결손)	있음	범주형																																																																																																																					
		age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target																																																																																																										
	87	53	0	3	128	216	0	2	115	0	0.0	1	0.0	NaN	0																																																																																																										
	166	52	1	3	138	223	0	0	169	0	0.0	1	NaN	3.0	0																																																																																																										
	192	43	1	4	132	247	1	2	143	1	0.1	2	NaN	7.0	1																																																																																																										
	266	52	1	4	128	204	1	0	156	1	1.0	2	0.0	NaN	1																																																																																																										
	287	58	1	2	125	220	0	0	144	0	0.4	2	NaN	7.0	0																																																																																																										
302	38	1	3	138	175	0	0	173	0	0.0	1	NaN	3.0	0																																																																																																											
				용어설명																																																																																																																					

과정명		PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명																		
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> <div>➤적용하기</div> <div>➤Outro</div> <div>•문제풀기</div>	<div>• 결측값이 있는 열의 값 분포: ca, thal</div> <div>df.ca.value_counts(dropna=False), df.thal.value_counts(dropna=False)</div> <div>↔ (ca</div> <table><tr><td>0.0</td><td>176</td></tr><tr><td>1.0</td><td>65</td></tr><tr><td>2.0</td><td>38</td></tr><tr><td>3.0</td><td>20</td></tr><tr><td>NaN</td><td>4</td></tr></table> <div>Name: count, dtype: int64,</div> <div>thal</div> <table><tr><td>3.0</td><td>166</td></tr><tr><td>7.0</td><td>117</td></tr><tr><td>6.0</td><td>18</td></tr><tr><td>NaN</td><td>2</td></tr></table> <div>Name: count, dtype: int64)</div>				0.0	176	1.0	65	2.0	38	3.0	20	NaN	4	3.0	166	7.0	117	6.0	18	NaN	2	
	0.0	176																					
	1.0	65																					
2.0	38																						
3.0	20																						
NaN	4																						
3.0	166																						
7.0	117																						
6.0	18																						
NaN	2																						
					용어설명																		

내레이션

15

과정명	PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> <div>➤적용하기</div> <div>➤Outro</div> <div>•문제풀기</div>	<div>• 결측값 처리</div> <div>연속형 자료 처리: 평균 값으로 대체</div> <div>범주형 자료 처리: 최빈 값으로 대체</div> <div>행 제거</div> <div><div>[19] # 연속형 'ca' 열의 결측값을 평균값으로 채움</div><div>df["ca"] = df["ca"].fillna(df["ca"].mean())</div><div># 범주형 'thal' 열의 결측값을 최빈값으로 채움</div><div>df["thal"] = df["thal"].fillna(df["thal"].mode()[0])</div><div>[20] # 하나라도 결측값이 있는 행이 없음을 확인</div><div>df[df.isna().any(axis=1)]</div><div><div>↺</div><div>age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal target</div><div></div></div></div>			
				용어설명

내레이션

16

과정명	PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명																																													
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤ 학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> <div>➤적용하기</div> <div>➤Outro</div> <div>•문제풀기</div>	<div>• 범주형 변수의 원-핫 인코딩</div> <div>범주형(categorical) 변수로서, 숫자로 표현되어 있지만 숫자 간에 크고 작음의 의미가 없음</div> <table><thead><tr><th>변수 이름</th><th>설명</th></tr></thead><tbody><tr><td>cp</td><td>가슴 통증 유형 (1~4)</td></tr><tr><td>restecg</td><td>심전도 결과 (0~2)</td></tr><tr><td>slope</td><td>ST 기울기 (1~3)</td></tr><tr><td>thal</td><td>탈라شم 검사 결과 (3, 6, 7)</td></tr></tbody></table> <div>함수 pd.get_dummies()</div> <div>범주형 변수를 각각의 고유 값(unique value) 별로 나누어 각각 하나의 열(column)로 바꿔 줌</div> <div>각 열은 해당 범주의 여부를 나타냄 (1: 해당, 0: 해당 아님)</div> <div>활용 예:</div> <div>'cp': [1, 2, 3, 4, 1, 2]</div> <table><thead><tr><th></th><th>cp_1</th><th>cp_2</th><th>cp_3</th><th>cp_4</th></tr></thead><tbody><tr><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>2</td><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>3</td><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>4</td><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>5</td><td>0</td><td>1</td><td>0</td><td>0</td></tr></tbody></table>			변수 이름	설명	cp	가슴 통증 유형 (1~4)	restecg	심전도 결과 (0~2)	slope	ST 기울기 (1~3)	thal	탈라شم 검사 결과 (3, 6, 7)		cp_1	cp_2	cp_3	cp_4	0	1	0	0	0	1	0	1	0	0	2	0	0	1	0	3	0	0	0	1	4	1	0	0	0	5	0	1	0	0	
	변수 이름	설명																																															
	cp	가슴 통증 유형 (1~4)																																															
	restecg	심전도 결과 (0~2)																																															
	slope	ST 기울기 (1~3)																																															
thal	탈라شم 검사 결과 (3, 6, 7)																																																
	cp_1	cp_2	cp_3	cp_4																																													
0	1	0	0	0																																													
1	0	1	0	0																																													
2	0	0	1	0																																													
3	0	0	0	1																																													
4	1	0	0	0																																													
5	0	1	0	0																																													
				용어설명																																													

과정명	PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명																																																																																																																																																												
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤ 학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> <div>➤적용하기</div> <div>➤ Outro</div> <div>•문제풀기</div>	<div>• 모든 범주형 변수를 원핫 인코딩으로 수정</div> <div>pd.get_dummies(df, columns=['cp'])[['cp_1', 'cp_2', 'cp_3', 'cp_4']]</div> <div>[37] pd.get_dummies(df, columns=['cp'])[['cp_1', 'cp_2', 'cp_3', 'cp_4']]</div> <div><div>↻</div><table><thead><tr><th></th><th>cp_1</th><th>cp_2</th><th>cp_3</th><th>cp_4</th></tr></thead><tbody><tr><td>0</td><td>True</td><td>False</td><td>False</td><td>False</td></tr><tr><td>1</td><td>False</td><td>False</td><td>False</td><td>True</td></tr><tr><td>2</td><td>False</td><td>False</td><td>False</td><td>True</td></tr><tr><td>3</td><td>False</td><td>False</td><td>True</td><td>False</td></tr><tr><td>4</td><td>False</td><td>True</td><td>False</td><td>False</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>298</td><td>True</td><td>False</td><td>False</td><td>False</td></tr><tr><td>299</td><td>False</td><td>False</td><td>False</td><td>True</td></tr><tr><td>300</td><td>False</td><td>False</td><td>False</td><td>True</td></tr><tr><td>301</td><td>False</td><td>True</td><td>False</td><td>False</td></tr><tr><td>302</td><td>False</td><td>False</td><td>True</td><td>False</td></tr></tbody></table><div>303 rows × 4 columns</div></div> <div><div>[39] # 모든 범주형 변수 원-핫 인코딩</div><div>categorical_cols = ["cp", "restecg", "slope", "thal"]</div><div>df = pd.get_dummies(df, columns=categorical_cols)</div><div>[41] df.info()</div><div><div>↻</div><div>&lt;class 'pandas.core.frame.DataFrame'&gt;</div><div>RangeIndex: 303 entries, 0 to 302</div><div>Data columns (total 23 columns):</div><table><thead><tr><th>#</th><th>Column</th><th>Non-Null Count</th><th>Dtype</th></tr></thead><tbody><tr><td>0</td><td>age</td><td>303 non-null</td><td>int64</td></tr><tr><td>1</td><td>sex</td><td>303 non-null</td><td>int64</td></tr><tr><td>2</td><td>trestbps</td><td>303 non-null</td><td>int64</td></tr><tr><td>3</td><td>chol</td><td>303 non-null</td><td>int64</td></tr><tr><td>4</td><td>fbs</td><td>303 non-null</td><td>int64</td></tr><tr><td>5</td><td>thalach</td><td>303 non-null</td><td>int64</td></tr><tr><td>6</td><td>exang</td><td>303 non-null</td><td>int64</td></tr><tr><td>7</td><td>oldpeak</td><td>303 non-null</td><td>float64</td></tr><tr><td>8</td><td>ca</td><td>303 non-null</td><td>float64</td></tr><tr><td>9</td><td>target</td><td>303 non-null</td><td>int64</td></tr><tr><td>10</td><td>cp_1</td><td>303 non-null</td><td>bool</td></tr><tr><td>11</td><td>cp_2</td><td>303 non-null</td><td>bool</td></tr><tr><td>12</td><td>cp_3</td><td>303 non-null</td><td>bool</td></tr><tr><td>13</td><td>cp_4</td><td>303 non-null</td><td>bool</td></tr><tr><td>14</td><td>restecg_0</td><td>303 non-null</td><td>bool</td></tr><tr><td>15</td><td>restecg_1</td><td>303 non-null</td><td>bool</td></tr><tr><td>16</td><td>restecg_2</td><td>303 non-null</td><td>bool</td></tr><tr><td>17</td><td>slope_1</td><td>303 non-null</td><td>bool</td></tr><tr><td>18</td><td>slope_2</td><td>303 non-null</td><td>bool</td></tr><tr><td>19</td><td>slope_3</td><td>303 non-null</td><td>bool</td></tr><tr><td>20</td><td>thal_3.0</td><td>303 non-null</td><td>bool</td></tr><tr><td>21</td><td>thal_6.0</td><td>303 non-null</td><td>bool</td></tr><tr><td>22</td><td>thal_7.0</td><td>303 non-null</td><td>bool</td></tr></tbody></table><div>dtypes: bool(13), float64(2), int64(8)</div><div>memory usage: 27.6 KB</div></div></div>				cp_1	cp_2	cp_3	cp_4	0	True	False	False	False	1	False	False	False	True	2	False	False	False	True	3	False	False	True	False	4	False	True	False	False	...	...	...	...	...	298	True	False	False	False	299	False	False	False	True	300	False	False	False	True	301	False	True	False	False	302	False	False	True	False	#	Column	Non-Null Count	Dtype	0	age	303 non-null	int64	1	sex	303 non-null	int64	2	trestbps	303 non-null	int64	3	chol	303 non-null	int64	4	fbs	303 non-null	int64	5	thalach	303 non-null	int64	6	exang	303 non-null	int64	7	oldpeak	303 non-null	float64	8	ca	303 non-null	float64	9	target	303 non-null	int64	10	cp_1	303 non-null	bool	11	cp_2	303 non-null	bool	12	cp_3	303 non-null	bool	13	cp_4	303 non-null	bool	14	restecg_0	303 non-null	bool	15	restecg_1	303 non-null	bool	16	restecg_2	303 non-null	bool	17	slope_1	303 non-null	bool	18	slope_2	303 non-null	bool	19	slope_3	303 non-null	bool	20	thal_3.0	303 non-null	bool	21	thal_6.0	303 non-null	bool	22	thal_7.0	303 non-null	bool	
	cp_1	cp_2	cp_3	cp_4																																																																																																																																																												
0	True	False	False	False																																																																																																																																																												
1	False	False	False	True																																																																																																																																																												
2	False	False	False	True																																																																																																																																																												
3	False	False	True	False																																																																																																																																																												
4	False	True	False	False																																																																																																																																																												
...	...	...	...	...																																																																																																																																																												
298	True	False	False	False																																																																																																																																																												
299	False	False	False	True																																																																																																																																																												
300	False	False	False	True																																																																																																																																																												
301	False	True	False	False																																																																																																																																																												
302	False	False	True	False																																																																																																																																																												
#	Column	Non-Null Count	Dtype																																																																																																																																																													
0	age	303 non-null	int64																																																																																																																																																													
1	sex	303 non-null	int64																																																																																																																																																													
2	trestbps	303 non-null	int64																																																																																																																																																													
3	chol	303 non-null	int64																																																																																																																																																													
4	fbs	303 non-null	int64																																																																																																																																																													
5	thalach	303 non-null	int64																																																																																																																																																													
6	exang	303 non-null	int64																																																																																																																																																													
7	oldpeak	303 non-null	float64																																																																																																																																																													
8	ca	303 non-null	float64																																																																																																																																																													
9	target	303 non-null	int64																																																																																																																																																													
10	cp_1	303 non-null	bool																																																																																																																																																													
11	cp_2	303 non-null	bool																																																																																																																																																													
12	cp_3	303 non-null	bool																																																																																																																																																													
13	cp_4	303 non-null	bool																																																																																																																																																													
14	restecg_0	303 non-null	bool																																																																																																																																																													
15	restecg_1	303 non-null	bool																																																																																																																																																													
16	restecg_2	303 non-null	bool																																																																																																																																																													
17	slope_1	303 non-null	bool																																																																																																																																																													
18	slope_2	303 non-null	bool																																																																																																																																																													
19	slope_3	303 non-null	bool																																																																																																																																																													
20	thal_3.0	303 non-null	bool																																																																																																																																																													
21	thal_6.0	303 non-null	bool																																																																																																																																																													
22	thal_7.0	303 non-null	bool																																																																																																																																																													
				용어설명																																																																																																																																																												

내레이션을

18



과정명		PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명
<div> <div>➤Intro</div> <div> <div>•학습열기</div> <div>•학습목표</div> </div> <div>➤ 학습하기</div> <div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> </div> <div>➤ 적용하기</div> <div>➤ Outro</div> <div>•문제풀기</div> </div>	<div> <div>간지</div> <div>데이터 정규화 및 시각화</div> </div>				
					용어설명
내레이션					19

과정명		PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명
<div> <div>➤Intro</div> <div> <div>•학습열기</div> <div>•학습목표</div> </div> <div>➤ 학습하기</div> <div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> </div> <div>➤ 적용하기</div> <div>➤ Outro</div> <div>•문제풀기</div> </div>	<div> <div>• 연속형 데이터의 정규화</div> <div>정규화는 특성(feature) 들의 값을 공통된 척도(scale)로 바꾸는 작업</div> <div>즉, 각 열의 값들을 0~1 범위 또는 평균 0, 표준편차 1 범위로 바꿈</div> <div>모든 특성이 동일한 비중으로 처리되도록 만드는 기법</div> <div>데이터들의 출발선을 똑같이 맞춰주는 일종의 체급 통일 작업</div> </div>				
					용어설명
내레이션					20

과정명	PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤ 학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> <div>➤적용하기</div> <div>➤ Outro</div> <div>•문제풀기</div>	<div>• 정규화의 필요성</div> <div>특성 간 단위 차이로 인한 학습 불균형 방지</div> <div>예: age: 29 ~ 77, chol: 100 ~ 564, oldpeak: 0.0 ~ 6.2</div> <div>→ 숫자가 큰 특성이 모델에서 더 중요한 것처럼 오해될 수 있음</div> <div>정규화 없이 거리 계산 시 큰 수치를 가진 특성이 결정적 영향을 줌</div> <div>경사하강법(Gradient Descent)의 수렴 속도 향상</div> <div>정규화를 하지 않으면 손실 함수의 모양이 비대칭해져서</div> <div>수렴 속도가 느려지거나 발산 위험이 있음</div> <div>→ 신경망, 로지스틱 회귀에서 매우 중요</div> <div>모델 성능 향상 및 안정성</div> <div>정규화는 과적합을 줄이고, 일관된 모델 성능을 만드는 데 도움 됨</div>			용어설명

내레이션

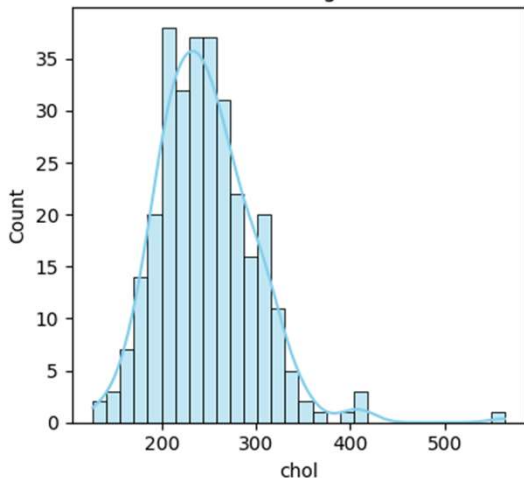
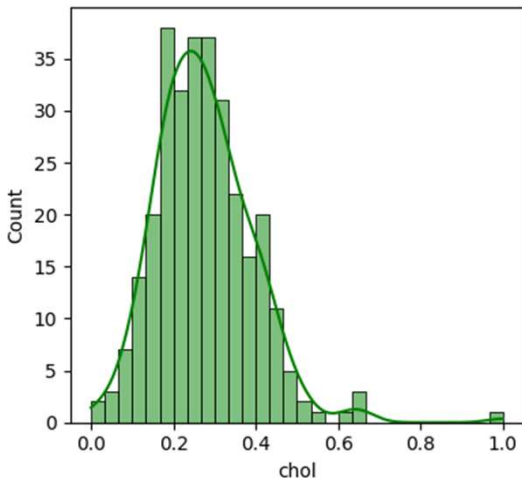
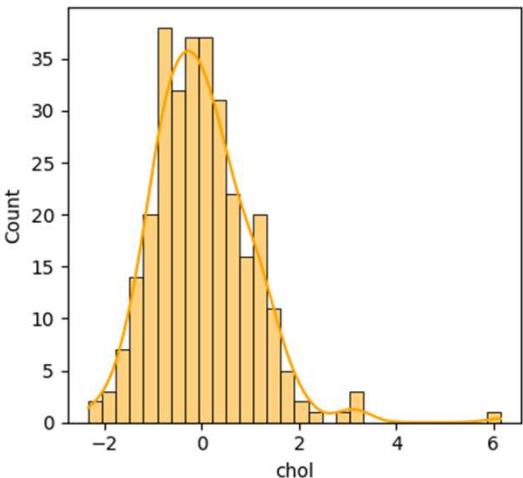
21

과정명		PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명															
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤ 학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> <div>➤ 적용하기</div> <div>➤ Outro</div> <div>•문제풀기</div>	<div>• 주요 정규화 기법</div> <table><thead><tr><th>정규화 기법</th><th>수식</th><th>특징</th></tr></thead><tbody><tr><td>Min-Max Scaling (최소-최대 정규화)</td><td><math>(x - \min) / (\max - \min)</math></td><td>0~1 사이로 조정, 간단함</td></tr><tr><td>Z-score Standardization (표준 정규화)</td><td><math>(x - \text{평균}) / \text{표준편차}</math></td><td>평균 0, 표준편차 1 → 정규분포 기반</td></tr><tr><td>Robust Scaling</td><td><math>(x - \text{median}) / \text{IQR}</math></td><td>이상치에 강함</td></tr><tr><td>MaxAbs Scaling</td><td><math>x / \max(\text{abs}(x))</math></td><td>희소 행렬에 적합</td></tr></tbody></table>				정규화 기법	수식	특징	Min-Max Scaling (최소-최대 정규화)	$(x - \min) / (\max - \min)$	0~1 사이로 조정, 간단함	Z-score Standardization (표준 정규화)	$(x - \text{평균}) / \text{표준편차}$	평균 0, 표준편차 1 → 정규분포 기반	Robust Scaling	$(x - \text{median}) / \text{IQR}$	이상치에 강함	MaxAbs Scaling	$x / \max(\text{abs}(x))$	희소 행렬에 적합	
	정규화 기법	수식	특징																	
	Min-Max Scaling (최소-최대 정규화)	$(x - \min) / (\max - \min)$	0~1 사이로 조정, 간단함																	
	Z-score Standardization (표준 정규화)	$(x - \text{평균}) / \text{표준편차}$	평균 0, 표준편차 1 → 정규분포 기반																	
	Robust Scaling	$(x - \text{median}) / \text{IQR}$	이상치에 강함																	
	MaxAbs Scaling	$x / \max(\text{abs}(x))$	희소 행렬에 적합																	
					용어설명															
내레이션					22															

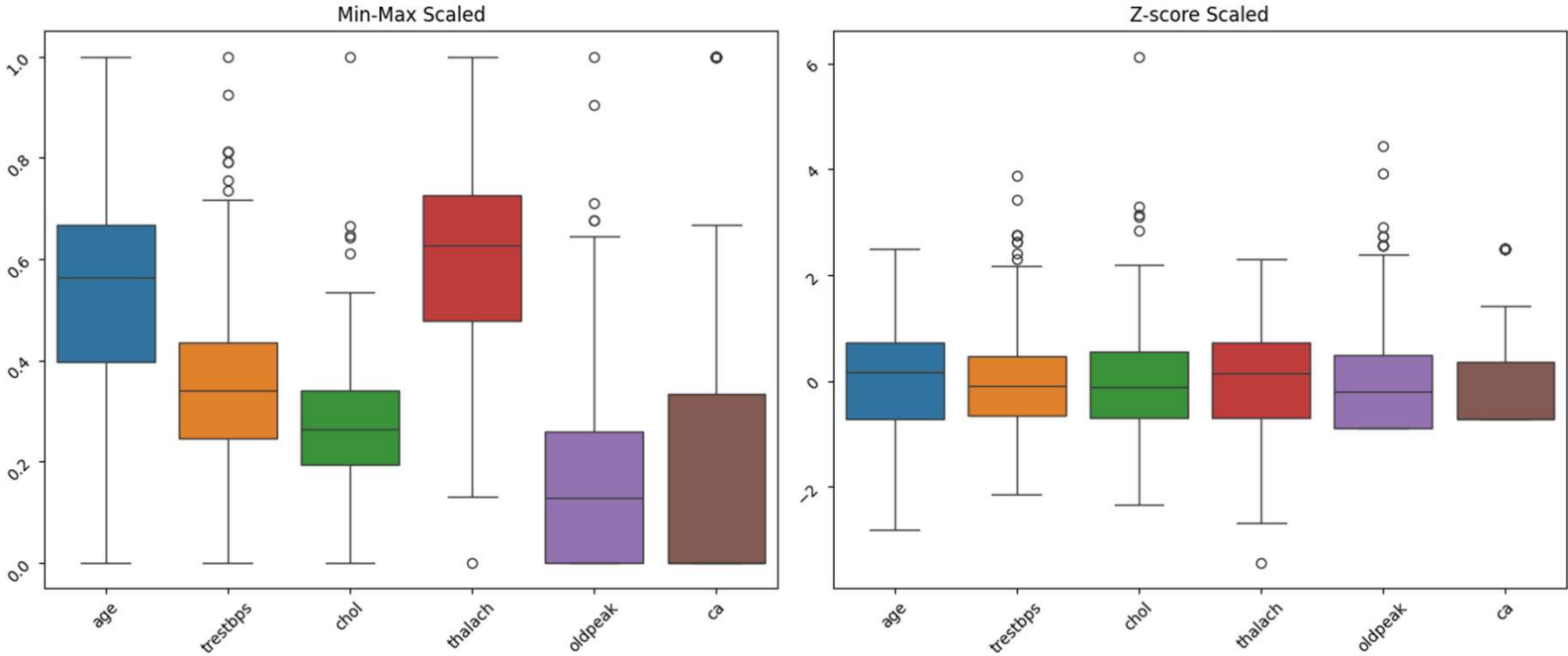
과정명		PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div>	<div>• 정규화 코드 준비</div> <div># 최소-최대 정규화, 표준 정규화 함수 사용</div> <div>from sklearn.preprocessing import MinMaxScaler, StandardScaler</div> <div># 정규화 대상: 연속형 수치 변수만 선택</div> <div>numeric_cols = ["age", "trestbps", "chol", "thalach", "oldpeak", "ca"]</div>				
					용어설명
<div>➤적용하기</div> <div>➤Outro</div> <div>•문제풀기</div>					
내레이션	23				

과정명		PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명																																																																																				
과정명	➤Intro	<div>• Min-Max Scaling</div> <div>모든 값을 0~1 사이로 변환 → 값의 상대적 크기 유지</div> <div><pre>[27] # Min-Max 정규화 minmax_scaler = MinMaxScaler() df_minmax = df.copy() df_minmax[numeric_cols] = minmax_scaler.fit_transform(df_minmax[numeric_cols])</pre></div> <div><pre>[28] df_minmax[numeric_cols]</pre></div> <div><div>↻</div><table><thead><tr><th></th><th>age</th><th>trestbps</th><th>chol</th><th>thalach</th><th>oldpeak</th><th>ca</th></tr></thead><tbody><tr><td>0</td><td>0.708333</td><td>0.481132</td><td>0.244292</td><td>0.603053</td><td>0.370968</td><td>0.000000</td></tr><tr><td>1</td><td>0.791667</td><td>0.622642</td><td>0.365297</td><td>0.282443</td><td>0.241935</td><td>1.000000</td></tr><tr><td>2</td><td>0.791667</td><td>0.245283</td><td>0.235160</td><td>0.442748</td><td>0.419355</td><td>0.666667</td></tr><tr><td>3</td><td>0.166667</td><td>0.339623</td><td>0.283105</td><td>0.885496</td><td>0.564516</td><td>0.000000</td></tr><tr><td>4</td><td>0.250000</td><td>0.339623</td><td>0.178082</td><td>0.770992</td><td>0.225806</td><td>0.000000</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>298</td><td>0.333333</td><td>0.150943</td><td>0.315068</td><td>0.465649</td><td>0.193548</td><td>0.000000</td></tr><tr><td>299</td><td>0.812500</td><td>0.471698</td><td>0.152968</td><td>0.534351</td><td>0.548387</td><td>0.666667</td></tr><tr><td>300</td><td>0.583333</td><td>0.339623</td><td>0.011416</td><td>0.335878</td><td>0.193548</td><td>0.333333</td></tr><tr><td>301</td><td>0.583333</td><td>0.339623</td><td>0.251142</td><td>0.786260</td><td>0.000000</td><td>0.333333</td></tr><tr><td>302</td><td>0.187500</td><td>0.415094</td><td>0.111872</td><td>0.778626</td><td>0.000000</td><td>0.224080</td></tr></tbody></table><div>303 rows × 6 columns</div></div>				age	trestbps	chol	thalach	oldpeak	ca	0	0.708333	0.481132	0.244292	0.603053	0.370968	0.000000	1	0.791667	0.622642	0.365297	0.282443	0.241935	1.000000	2	0.791667	0.245283	0.235160	0.442748	0.419355	0.666667	3	0.166667	0.339623	0.283105	0.885496	0.564516	0.000000	4	0.250000	0.339623	0.178082	0.770992	0.225806	0.000000	...	...	...	...	...	...	...	298	0.333333	0.150943	0.315068	0.465649	0.193548	0.000000	299	0.812500	0.471698	0.152968	0.534351	0.548387	0.666667	300	0.583333	0.339623	0.011416	0.335878	0.193548	0.333333	301	0.583333	0.339623	0.251142	0.786260	0.000000	0.333333	302	0.187500	0.415094	0.111872	0.778626	0.000000	0.224080	
					age	trestbps	chol	thalach	oldpeak	ca																																																																															
	0				0.708333	0.481132	0.244292	0.603053	0.370968	0.000000																																																																															
1	0.791667	0.622642	0.365297	0.282443	0.241935	1.000000																																																																																			
2	0.791667	0.245283	0.235160	0.442748	0.419355	0.666667																																																																																			
3	0.166667	0.339623	0.283105	0.885496	0.564516	0.000000																																																																																			
4	0.250000	0.339623	0.178082	0.770992	0.225806	0.000000																																																																																			
...	...	...	...	...	...	...																																																																																			
298	0.333333	0.150943	0.315068	0.465649	0.193548	0.000000																																																																																			
299	0.812500	0.471698	0.152968	0.534351	0.548387	0.666667																																																																																			
300	0.583333	0.339623	0.011416	0.335878	0.193548	0.333333																																																																																			
301	0.583333	0.339623	0.251142	0.786260	0.000000	0.333333																																																																																			
302	0.187500	0.415094	0.111872	0.778626	0.000000	0.224080																																																																																			
	➤학습하기																																																																																								
	1. 코딩 환경 구축																																																																																								
	2. 심장병 데이터 전처리																																																																																								
	3. 데이터 정규화 및 시각화																																																																																								
	➤적용하기																																																																																								
	➤Outro																																																																																								
	•문제풀기																																																																																								
내레이션																																																																																									

과정명		PyTorch로 배우는 머신러닝 알고리즘		회차명	4	화면설명																																																																																				
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤ 학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div>	<div>• Z-score Scaling: 평균 0, 표준편차 1</div> <div>이상치 감지, 정규분포 기반 모델에 적합</div> <div><div>[29] # Z-score 정규화</div><div>standard_scaler = StandardScaler()</div><div>df_zscore = df.copy()</div><div>df_zscore[numeric_cols] = standard_scaler.fit_transform(df_zscore[numeric_cols])</div></div> <div><div>▶ df_zscore[numeric_cols]</div><div><div>↻</div><table><thead><tr><th></th><th>age</th><th>trestbps</th><th>chol</th><th>thalach</th><th>oldpeak</th><th>ca</th></tr></thead><tbody><tr><td>0</td><td>0.948726</td><td>0.757525</td><td>-0.264900</td><td>0.017197</td><td>1.087338</td><td>-0.723095</td></tr><tr><td>1</td><td>1.392002</td><td>1.611220</td><td>0.760415</td><td>-1.821905</td><td>0.397182</td><td>2.503851</td></tr><tr><td>2</td><td>1.392002</td><td>-0.665300</td><td>-0.342283</td><td>-0.902354</td><td>1.346147</td><td>1.428203</td></tr><tr><td>3</td><td>-1.932564</td><td>-0.096170</td><td>0.063974</td><td>1.637359</td><td>2.122573</td><td>-0.723095</td></tr><tr><td>4</td><td>-1.489288</td><td>-0.096170</td><td>-0.825922</td><td>0.980537</td><td>0.310912</td><td>-0.723095</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>298</td><td>-1.046013</td><td>-1.234430</td><td>0.334813</td><td>-0.770990</td><td>0.138373</td><td>-0.723095</td></tr><tr><td>299</td><td>1.502821</td><td>0.700612</td><td>-1.038723</td><td>-0.376896</td><td>2.036303</td><td>1.428203</td></tr><tr><td>300</td><td>0.283813</td><td>-0.096170</td><td>-2.238149</td><td>-1.515388</td><td>0.138373</td><td>0.352554</td></tr><tr><td>301</td><td>0.283813</td><td>-0.096170</td><td>-0.206864</td><td>1.068113</td><td>-0.896862</td><td>0.352554</td></tr><tr><td>302</td><td>-1.821745</td><td>0.359134</td><td>-1.386944</td><td>1.024325</td><td>-0.896862</td><td>0.000000</td></tr></tbody></table><div>303 rows × 6 columns</div></div></div>						age	trestbps	chol	thalach	oldpeak	ca	0	0.948726	0.757525	-0.264900	0.017197	1.087338	-0.723095	1	1.392002	1.611220	0.760415	-1.821905	0.397182	2.503851	2	1.392002	-0.665300	-0.342283	-0.902354	1.346147	1.428203	3	-1.932564	-0.096170	0.063974	1.637359	2.122573	-0.723095	4	-1.489288	-0.096170	-0.825922	0.980537	0.310912	-0.723095	...	...	...	...	...	...	...	298	-1.046013	-1.234430	0.334813	-0.770990	0.138373	-0.723095	299	1.502821	0.700612	-1.038723	-0.376896	2.036303	1.428203	300	0.283813	-0.096170	-2.238149	-1.515388	0.138373	0.352554	301	0.283813	-0.096170	-0.206864	1.068113	-0.896862	0.352554	302	-1.821745	0.359134	-1.386944	1.024325	-0.896862	0.000000	
		age	trestbps	chol	thalach	oldpeak	ca																																																																																			
	0	0.948726	0.757525	-0.264900	0.017197	1.087338	-0.723095																																																																																			
1	1.392002	1.611220	0.760415	-1.821905	0.397182	2.503851																																																																																				
2	1.392002	-0.665300	-0.342283	-0.902354	1.346147	1.428203																																																																																				
3	-1.932564	-0.096170	0.063974	1.637359	2.122573	-0.723095																																																																																				
4	-1.489288	-0.096170	-0.825922	0.980537	0.310912	-0.723095																																																																																				
...	...	...	...	...	...	...																																																																																				
298	-1.046013	-1.234430	0.334813	-0.770990	0.138373	-0.723095																																																																																				
299	1.502821	0.700612	-1.038723	-0.376896	2.036303	1.428203																																																																																				
300	0.283813	-0.096170	-2.238149	-1.515388	0.138373	0.352554																																																																																				
301	0.283813	-0.096170	-0.206864	1.068113	-0.896862	0.352554																																																																																				
302	-1.821745	0.359134	-1.386944	1.024325	-0.896862	0.000000																																																																																				
<div>➤적용하기</div> <div>➤ Outro</div> <div>•문제풀기</div>						용어설명																																																																																				
내레이션						25																																																																																				

과정명	PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤ 학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> <div>➤ 적용하기</div> <div>➤ Outro</div> <div>•문제풀기</div>	<div>• dots 분포(histplot)로 연속형 변수 분포 비교</div> <div>정규화 이전, 최소최대 정규화, 표준정규화</div> <div><div>chol - Original</div></div> <div><div>chol - MinMax Scaled</div></div> <div><div>chol - Zscore Scaled</div></div>			
내레이션				<div>용어설명</div>
				26
				2



과정명	PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명
<div>➤Intro</div> <ul style="list-style-type: none"> <li>•학습열기</li> <li>•학습목표</li> </ul> <div>➤ 학습하기</div> <ol style="list-style-type: none"> <li>1. 코딩 환경 구축</li> <li>2. 심장병 데이터 전처리</li> <li>3. 데이터 정규화 및 시각화</li> </ol> <div>➤ 적용하기</div> <div>➤ Outro</div> <ul style="list-style-type: none"> <li>•문제풀기</li> </ul>	<div> <div>• 두 정규화 기법: 박스 플롯으로 연속형 변수 분포 비교</div> <div>  </div> </div> <div data-bbox="1912 236 2228 1225"></div>			용어설명
	27			2

과정명	PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명
<div>➤Intro</div> <div>•학습열기</div> <div>•학습목표</div> <div>➤ 학습하기</div> <div>1. 코딩 환경 구축</div> <div>2. 심장병 데이터 전처리</div> <div>3. 데이터 정규화 및 시각화</div> <div>➤ 적용하기</div> <div>➤ Outro</div> <div>•문제풀기</div>	<div>• 전처리한 데이터셋 파일로 저장</div> <div>다음 시간 DNN 구현에서 파일 사용: heart_disease_zscore.csv</div> <div># 원-핫 인코딩 전처리한 파일</div> <div>df.to_csv("heart_disease_preprocessed.csv", index=False)</div> <div># 원-핫 인코딩, 최소-최대 정규화 전처리한 파일</div> <div>df_minmax.to_csv("heart_disease_minmax.csv", index=False)</div> <div># 원-핫 인코딩, 표준 정규화 전처리한 파일, 이 파일을 다음 시간에 사용</div> <div>df_zscore.to_csv("heart_disease_zscore.csv", index=False)</div>			
				용어설명
내레이션	28			

과정명		PyTorch로 배우는 머신러닝 알고리즘	회차명	4	화면설명
과정명	➤Intro	• 학습열기 • 학습목표	➤ 학습하기	1. 코딩 환경 구축 2. 심장병 데이터 전처리 3. 데이터 정규화 및 시각화	① 학습 내용과 관련하여 실제 적용력을 높일 수 있는 문제, 혹은 주제를 작성해 주세요.  ② ex. 사례 제시 후 전문가 의견, 실습과제, 응용 예시 시뮬레이션 등  ③ 저작권 침해가 되지 않도록 내용을 구성해 주세요.  ④ 출처가 있을 경우 반드시 작성해 주세요.
	➤적용하기				
		PyTorch로 배우는 머신러닝 알고리즘			화면설명
		회차명			4
		Min-Max Scaling은 데이터를 0과 1 사이로 스케일링하여 값의 상대적인 크기를 보존합니다. 구현이 간단하고, 값의 범위가 고정된 경우 적절합니다. 하지만 이상치에 매우 민감하다는 단점이 있습니다. 반면, Z-score 정규화는 평균이 0이고 표준편차가 1이 되도록 변환하며, 정규분포를 가정한 알고리즘(예: 선형 회귀, 로지스틱 회귀, 신경망 등)에 유리합니다. 이상치가 존재해도 상대적으로 영향을 덜 받습니다. 따라서 데이터에 이상치가 많다면 Z-score 방식이 더 적합하며, 이상치가 거의 없고 값을 0~1 범위로 고정해야 하는 상황에서는 Min-Max Scaling이 유리할 수 있습니다. 정규화는 모델의 수렴 속도를 높이고, 경사하강법 기반 학습에서 손실 함수의 왜곡을 방지하여 더 안정적인 학습을 가능하게 합니다			
		용어설명			
		29			