

AI 알고리즘&함수 정리

The background of the slide is an abstract composition of large, overlapping brushstrokes. The top half is dominated by a bright yellow stroke. Below this, there are broad strokes of light blue and a soft pink. The bottom right corner features a darker blue stroke. The overall effect is a modern, artistic, and colorful backdrop for the title.

번호	구분	알고리즘명 (함수명)	의미 & 사용 시점	사용법 (메서드)
1	머신러닝 (지도학습)	k-최근접이웃 알고리즘	가장 가까운 k개의 데이터를 보고 다중결의 원칙에 따라 데이터를 예측/분류	<code>KNeighborsClassifier()</code> <code>fit()</code> / <code>score()</code> / <code>predict()</code>
2	numpy	<code>np.arange(n)</code>	0~n까지 1씩 증가하는 인덱스 생성	
3	numpy	<code>np.random.shuffle(t)</code>	t 배열을 무작위로 섞음.	
4	numpy	<code>np.column_stack([l1, l2, l3])</code> <code>[l1, l2, l3])</code>	([[1, 4, 2, 5, 3, 6]]) 리스트를 세는 뒤 행으로 묶임.	
5	numpy	<code>np.concatenate([l1, l2, l3])</code> <code>[l1, l2, l3])</code>	[[1, 1, 1, 0, 0, 0]]처럼 배열을 옆으로 붙여줌	
6	numpy	<code>np.ones(n)</code> <code>np.zeros(n)</code>	n만큼 1로 채워진 배열 n만큼 0으로 채워진 배열	
7	sklearn	<code>train_test_split()</code>	test set과 train set 분리	<code>train_input, test_input, test_target,</code> <code>train_target = train_test_split</code> <code>(a=target, a=data)</code>
8	numpy	<code>np.mean(t, axis=0)</code>	t 배열의 평균을 행을 따라 계산	
9	numpy	<code>np.std(t, axis=1)</code>	t 배열의 표준편차를 열을 따라 계산	

번호	구분	알고리즘명 (함명)	의미 & 사용 시점	사용법
10	머신러닝 (지도학습)	k-최근접 이웃 회귀	예측하려는 샘플에 가장 가까운 샘플 k개 (임의의 수치)의 평균을 예측 타겟값으로 함	KNeighborsRegressor (n-neighbors = k)
11	numpy	<code>t = t.reshape(a,b)</code>	t 배열을 a,b 크기로 바꿔줌	<code>[1,2,3,4]</code> → <code>[1,2],[3,4]</code>
12	sklearn	<code>mean_absolute_error(t, p)</code>	테스트 케이스 t에 대한 예측 p의 절대값오차 반환	
13	머신러닝 (지도학습)	선형 회귀	특성이 1개인 경우 어떤 직선을 학습 하는 알고리즘	<code>linear regression()</code>
14	Linear Regression	<code>lr.coef_</code>	직선의 기울기	
15	Linear Regression	<code>lr.intercept_</code>	직선의 절편	
16	pandas	<code>pd.read_csv(경로)</code>	경로에 위치한 CSV파일 읽기	
17	numpy	<code>np.array([a,b,c])</code>	배열 생성	
18	sklearn	<code>Polynomial Features()</code>	변환기 (전처리 위한 클래스)	<code>fit(t)</code> → 샘플 t 훈련 <code>transform(t)</code> → 샘플별 t 변환

번호	구분	알고리즘명 (함수명)	의미 or 사용 시점	사용법
19	sklearn	Standard Scaler()	특성을 표준점수로 바꾸는 변환기 (전처리 위한 클래스)	fit, transform
20	sklearn	Ridge(alpha)	릿지 회귀 (선형 회귀모델의 규제)	계수를 제곱한 값을 기반으로 규제 적용
21	sklearn	Lasso(alpha)	라쏘 회귀 (선형 회귀모델의 규제)	계수의 절댓값을 기반으로 규제 적용
22	pandas	pd.unique(a['b'])	a 데이터 프레임의 b열에서 고유한 값 추출	
23	numpy	a['b'].to_numpy()	넘파이 배열로 변환	
24	sklearn	predict_proba	클래스별 확률값 반환	
25	매트러닝 (지도학습)	로지스틱 회귀	선형 회귀와 동일하게 방정식을 학습 하지만 분류 모델	Logistic Regression()
26	Scipy	expit(n)	Sigmoid 함수 (하나의 방정식의 출력 값을 0~1로 압축)	
27	Scipy	Softmax(n, axis=1)	샘플 n에 속함 이기 대해 softmax 계산 (여러 방정식의 출력 값을 0~1로 압축, 전체 합이 1)	

번호	구분	알고리즘명 (함수명)	의미 or 사용 시점	사용법
28	sklearn	SGDClassifier	경사 하강법	SGDClassifier(loss='log', max_iter=10, random_state=42) ↓ 32비트 정수 ↓ 반복 에포크 횟수
29	sklearn	a.partial_fit(b, c)	경사 하강법의 추가 훈련	
30	pandas	a.head(c)	데이터 프레임 a의 첫 5개 샘플 출력	
31	pandas	a.info()	데이터 프레임 a의 각 열의 데이터 타입과 누락된 정보 표시	
32	pandas	a.describe()	데이터 프레임 a의 간략한 통계(최소, 최대, 평균값) 출력	
33	sklearn	DecisionTreeClassifier	결정 트리 모델	DecisionTreeClassifier(random_state=42, max_depth=3) → 루트 노드 아래 최대 3개 노드 출력
34	sklearn	plot_tree(a)	결정 트리 모델 a의 시각화	
35	결정트리	dt.feature_importances_	어떤 특성이 유용한지 나타내는 특성 중요도 출력	
36	sklearn	교차검증 함수	안정적인 검증 결과 획득, 훈련에 더 많은 데이터 사용	CrossValidate(a, input, target) ↓ 평가할 객체 ↓ 훈련 세트 전체

번호	구분	알고리즘명 (함수명)	의미 or 사용 시점	사용법
37	sklearn	StratifiedKFold	교차 검증시 훈련세트를 섞는 분할기의 일종.	<code>cross_validate(a, input, target, cv=StratifiedKFold())</code>
38	sklearn	Grid SearchCV	하이퍼 파라미터 탐색 + 교차검증	
39	numpy	np.argmax	가장 큰 값의 인덱스 추출	
40	sklearn	Random Forest Classifier	정형 데이터를 군다루는 앙상블 학습의 대표 주자 중 하나.	<code>RandomForestClassifier(n_jobs=-1)</code> CPU 코어 수 지정 시 -1 일 경우 모두 사용
41	sklearn	Extra Trees Classifier	과대적합 방지, 검증 세트 정수 향상	
42	sklearn	Gradient Boosting Classifier	깊이가 얕은 결정 트리 사용 → 이전 트리의 오차 보완	
43	sklearn	<code>enable_hist_gradient_boosting</code>	입력 특성을 256개 구간으로 나눠 보드 분할시 최적의 분할을 빠르게 탐색	
44	sklearn	<code>Permutation_importance</code>	특성을 하나씩 간헐하게 섞기 모델 성능 변화 관찰 → 특성 계량	
45	xgboost	<code>XGBClassifier</code>	히스토그램 기반 그래디언트 부스팅 알고리즘.	<code>XGBClassifier(tree_method='hist' ...)</code> hist 설정 시 히스토그램 기반 그래디언트 부스팅

번호	구분	알고리즘명 (함수명)	의미 or 사용 시점	사용법
46	lightgbm	LGBMClassifier	히스토그램 기반 그래디언트 부스팅 알고리즘.	
47	numpy	np.load ('파일명')	numpy에서 npy 파일 로드	
48	비지도학습 sklearn	k-평균 알고리즘.	k개의 클러스터 중심을 정한 뒤, 샘플의 표함값으로 클러스터 중심을 변경.	KMeans (n_clusters=3, ...) ↳ 클러스터 개수 지정
49	KMeans	km.labels_	구분된 결과 출력	
50	KMeans	km.inertia_	클러스터 중심과 클러스터에 속한 샘플 사이의 거리를 재는 것	
51	sklearn	PCA	데이터를 잘 나타내는 일부 특성을 선택해 차원을 축소하는 것.	PCA(n_components=50) ↳ 주성분 개수
52	pca	pca.inverse_transform(a)	축소된 a의 원본 데이터를 재구성	
53	pca	pca.explained_variance_ratio_	각 주성분의 설명된 분산 비율 출력	
54	tensor flow	keras	keras 패키지	(train_input, train_target), (test_input, test_target) = \keras. datasets.fashion_mnist.load_data()

번호	구분	알고리즘명 (함수명)	의미 or 사용 시점	사용법
55	keras	dense	밀접층 구현 시 필요 함수 (이진분류: sigmoid, 다중분류: Softmax)	keras.layers.Dense (10, activation='softmax', input_shape=(784,)) ↓ 노출력 크기 노란색 노란색 노란색
56	keras	keras.Sequential(dense)	dense로 만든 밀접층으로 신경망 모델 구현.	
57	keras	compile	케라스 모델의 훈련 전 설정. 손실함수, 계산하고픈 측정값 지정 (이진분류: 'binary_crossentropy', 다중분류: 'categorical_crossentropy')	t.compile(loss='sparse_categorical_crossentropy', metrics=['accuracy'])
58	keras	evaluate	케라스에서 모델 성능 평가 시 사용	
59	keras	t.add()	입력층 바로 뒤에 층 추가	t.add(keras.layers.Flatten(input_shape=(28, 28)))
60	keras	optimizer	케라스의 경사 하강법 알고리즘	keras.optimizers.SGD (learning_rate=0.1) ↓ 학습률
61	keras	비테로프 모델링 최적화	경사 하강법 옵티마이저	keras.optimizers.SGD (momentum=0.9, nesterov=True)
62	keras	Adagrad()	모델이 최적점에 갈수록 학습률 낮추는 적응적 학습률 옵티마이저	keras.optimizers.Adagrad() t.compile(optimizer=adagrad...)
63	keras	RMSprop()	모델이 최적점에 갈수록 학습률 낮추는 적응적 학습률 옵티마이저	keras.optimizers.RMSprop() t.compile(optimizer=RMSprop...)

번호	구분	알고리즘명 (함수명)	의미 or 사용 시점	사용법
64	keras	Adam()	모델이 최적점에 갈수록 학습률 낮추는 적응적 학습률 옵티마이저 Adam : 모멘텀 최적화와 RMSProp의 장점을 결합	<code>t.compile(optimizer = adam ...</code>
65	keras	<code>keras.layers.Dropout(0.3)</code>	드롭아웃 : 일부 뉴런을 랜덤하게 꺼서 과대적합 방지	
66	keras	<code>t.save_weights('name')</code> <code>t.save('name.h5')</code>	각각 훈련된 케라스 모델의 파라미터 저장과 모델 구조 + 파라미터 저장.	
67	keras	<code>t.load_weights('name.h5')</code>	이전에 저장한 모델 적재	
68	keras	Model Check point callback	keras에서 모델 과대적합 전에 막는 조기종료	<code>keras.callbacks.ModelCheckpoint('name')</code>
69	keras	Early Stopping Callback	Model Check point callback과 함께 쓸 시 검증 손실에 따라 훈련 중지 및 최상 파라미터로 복구	<code>keras.callbacks.EarlyStopping(patience=2, restore_best_weights=True)</code>
70	keras	<code>keras.layers.Conv2D</code>	입력 위를 이동하는 합성곱	<code>keras.layers.Conv2D(10, kernel_size=(3,3), activation='relu')</code>
71	keras	same padding (0번개변수)	이미지 주변 정보 보존을 위한 same padding ↔ valid padding	<code>keras.layers.Conv2D(... padding='same')</code>
72	keras	strides (0번개변수)	합성곱 연산시 이동하는 칸을 조절하여 특성 맵 크기 조절	<code>keras.layers.Conv2D(... strides=1)</code>

번호	구분	알고리즘명 (함수명)	의미 or 사용 시점	사용법
73	keras	pooling	특성 맵의 크기를 [최대값 기준으로 줄이는 최대 풀링 평균값 기준으로 줄이는 평균 풀링]	<code>keras.layers.MaxPooling2D(2, strides=2, padding='valid')</code>
74	keras	plot_model	합성 신경망 모델의 층의 구성을 그림으로 표현해 줌. (대개 변수: <code>show_shapes</code> → 그림의 입/출력 크기 표시 <code>to_file</code> → 파일로 저장 <code>dpi</code> → 해상도 지정)	<code>keras.utils.plot_model(model)</code>
75	keras	layers	캐라드 모델에 추가한 층 보기 (리스트로)	<code>model.layers</code>
76	keras	Model	여러 input 과 output을 연결	<code>model = keras.Model(inputs, outputs)</code>
77	keras	Input	Input Layer 클래스 객체를 만들어 출력 반환	<code>inputs = keras.Input(shape=(784,))</code>
78	keras	pad_sequences	시퀀스 데이터 길이를 맞춰줌. 원하는 길이보다 긴 경우 잘라내고 짧은 경우 0으로 패딩함.	<code>pad_sequences(train_input, maxlen=100)</code> → 원하는 길이
79	keras	to_categorical	원-핫 인코딩	<code>keras.utils.to_categorical(train_seq)</code>
80	keras	SimpleRNN	순환신경망	<code>keras.layers.SimpleRNN(8, input_shape=(100, 500))</code>
81	keras	Embedding	각 단어를 고정된 크기의 실수 벡터로 바꿔주는 embedding 기능	<code>model.add(keras.layers.Embedding(200, 16, input_length=100))</code> ↓ ↳ 임베딩 ↳ 임베딩 시퀀스 크기 사전 크기 벡터 크기

[illegible]