

분류	주요 내용	학습방법
1. 머신러닝 패키지	scikit-learn 구조 설명 scikit-learn 샘플데이터 불러오기, 만들기 - sklearn.datasets.xxx() - make_classification(), make_blobs(), make_circle(), make_moon() Iris 품종 분류 데이터 적재 및 정보확인 데이터 분할: train_test_split() DecisionTree 알고리즘 cross-fold(데이터 분할), cross_val_score() Hyperparameter 튜닝: GridSearch()	강의, 실습
2. 데이터 전처리	X값: 범주형-Encoding - Label Encoding - Ordinal Encoding - One-hot Encoding - Mean Encoding X값: 수치형-Scaling - Standard Scaling - Robust Scaling - MinMax Scaling - Normalization Scaling - Log Scaling - Exponential Scaling Feature Engineering - 구간 분할, 이산화 - 상호작용과 다항식 특성 - 특성 자동 선택 Missing value, Outlier 처리 언더 샘플링 / 오버 샘플링(SMOTE)	강의, 실습

분류	주요 내용	학습방법
3. 지도학습 평가	회귀 평가 - RMSE, MSE, ... 분류 평가 - Accuracy, Precision, Recall, F1-score, ROC, ...	강의, 실습
4. 선형 회귀	선형 회귀 알고리즘 설명 및 실행 경사하강법 Over fitting 규제: Ridge, Lasso, ElasticNet LogisticRegression (분류) - 알고리즘 설명 및 실행 - Sigmoid, Softmax SVM(Support Vector Machine) - 회귀와 분류의 차이점 인공 신경망 - 멀티레이어 퍼셉트론	강의, 실습
5. 결정 트리	Gini Index, Entropy Index Information Gain Feature Importance - Information Gain의 평균 - Permutation Feature Importance - 모든 모델에 적용 가능	강의, 실습
6. 앙상블	Voting, Bagging, Boosting, Stacking RandomForest - 회귀, 분류의 예측값 계산 방법(hard voting, soft voting) AdaBoost, GBM, XGBoost, LightGBM, CatBoost - 회귀: 잔차를 예측 - 분류: logloss 예측	강의, 실습

분류	주요 내용	학습방법
7. 차원 축소	PCA(Principal Component Analysis) - 주성분 분석 LDA(Linear Discriminant Analysis) - 선형 판별 분석 SVD(Singular Value Decomposition) - 특이값 분해 NMF(Non-Negative Matrix Factorization) - 비음수 행렬 분해 t-SNE(t-Distributed Stochastic Neighbor Embedding)	강의, 실습
8. 군집 분석	KMeans 병합군집(Agglomerative Clustering) GMM(Gaussian Mixture Model) DBSCAN 군집 평가: ARI, NM, Silhouette Coefficient	강의, 실습
9. 텍스트 분석	BOW(n-gram) 불용어 처리 TF-IDF 감성분석: IDMB 영화평(지도학습) 토픽모델링(LDA)	강의, 실습