

커리큘럼 _SeniorDS 이론

과목	소과목	상세내용	SeniorDS (Level2)	SeniorDS 강의일차
① 통계학	추정과 검정 기초	비모수, 다변량 검정(Wilcoxon test, Kruskal Wallis test, Friedman Test, Hotelling's T Square)	●	1
	Time series Analysis	Stationarity/Autocorrelation, Unit root test	●	
		Exponential Smoothing	●	
		- Double Exponential Smoothing(추세변동 고려; Holt법)	●	
		- Triple Exponential Smoothing(추세,계절변동 고려; HoltWinters)	●	
		Box-Jenkins ARIMA(Autoregressive Integrated Moving Average)	●	
② 데이터 처리 및 시각화	데이터 Cleansing	Outlier의 처리 - $x \sim y$ 관점에서의 제거 (회귀잔차)/다변량분포 기반	●	2
③ 머신러닝	러닝의 기초	머신러닝 방법론의 분류 및 전체적인 컨셉 - Supervised learning (Regression, Classification, Deep Learning) - Unsupervised learning(Clustering, Dimension Reduction, Recommendation) - Reinforcement learning	●	
		데이터셋의 분리 - Evaluate model : 모델 평가 지표(차트 포함) : Regression / Classification 모델 평가 measure (RMSE, MAE, MAPE...), (ROC, AUC, Lift, Recall, Precision...) - Evaluation of accuracy: Bias vs. Variance validation 방법(k-fold cross validation 등)	●	
	Clustering	Mean-shift clustering	●	
		DBSCAN	●	
		Gaussian Mixture Model	●	
	Feature Engineering, Feature Selection 및 Dimension Reduction	다중 t검정 기반	●	
		decision tree 기반	●	
		wrapper 기반 선택: local search, genetic algorithm heuristic	●	
		Principal Component Analysis(PCA)	●	
		t-Stochastic Nearest Neighbor(t-SNE)	●	
		Numerosity reduction - parametric methods - Non-parametric methods	●	
② 데이터 처리 및 시각화	Sampling	Classification의 클래스 불균형 데이터의 전처리 : 언더/오버/하이브리드 샘플링	●	3
③ 머신러닝	Regression	Multiple Linear Regression(변수선택, 가정의 검토, 다중공선성의 처리) - 가정 위반 확인법 및 해결책 - 변수선택(Forward, Backward, stepwise, R2 법 등...)	●	
		- 이상점, 영향점 관련 measure, 해결책 - 다중공선성 확인(VIF,상관계수 확인) 및 모형에 미치는 영향 - Regularization(Ridge/Lasso/Elastic Net)	●	
		SVM(Support Vector Machine)	●	
		Binary Classification & Multi-class Classification	●	
		Ensemble Methods : Bagging 방법 (Random Forest 등)	●	
	Tree Model	Ensemble Methods : Boosting 방법 (Gradient Boosting Tree 등)	●	
		Least square	●	
	머신러닝 관련 최적화	Gradient descent	●	
		EM algorithm	●	
	Recommendation	Contents based filtering	●	
		Collaborative filtering	●	
		Matrix Factorization과 ALS	●	
④ Deep Learning	Neural Neterwork 기본	Perceptron, Activation function 네트워크 최적화 (Activation function, initialization method, drop-out, batch normalization) 데이터 처리 기법(data augmentation) Multi-Layer Perceptron (parameters : epch, loss, optimizer) Backpropagation	●	4~5
	CNN	convolutional & pooling layer image CNN	●	
	RNN	recurent layer lstm, GRU time series or text(word2vec 기본)	●	
	기타 Deep Learning	비지도 : Auto Encoder(variational) Adversarial Model(GAN) Transformer 대표적인 network 소개(vgg, inception etc) Deep Learning 한계 : Adversarial	●	
	강화학습	기본 개념 및 예제 Q Learning	●	
⑤ 텍스트 분석	텍스트 분석 기본	텍스트 분석의 이해 및 자연어처리 기초	●	
	텍스트 전처리	StopWord Remove, Tokenize, Keyword 도출, 문장 요약	●	
	텍스트시각화	Word Cloud, 키워드 연관 네트워크 등	●	
	텍스트 머신러닝	Term feature weighting 방법 (TF-IDF, Normalization, ...)	●	
		확률론적 언어 모델링 방법, N-Gram	●	
		Conditional Random Fields (CRF) 기반 Named Entity Recognition (NER)	●	
		감성 분석(Sentiment analysis), 문장 요약, 유사도 분석	●	
		토픽 모델링을 위한 LDA	●	