2020 Fall: Business Analytics

Pilsung Kang

School of Industrial Management Engineering

Korea University

Overview

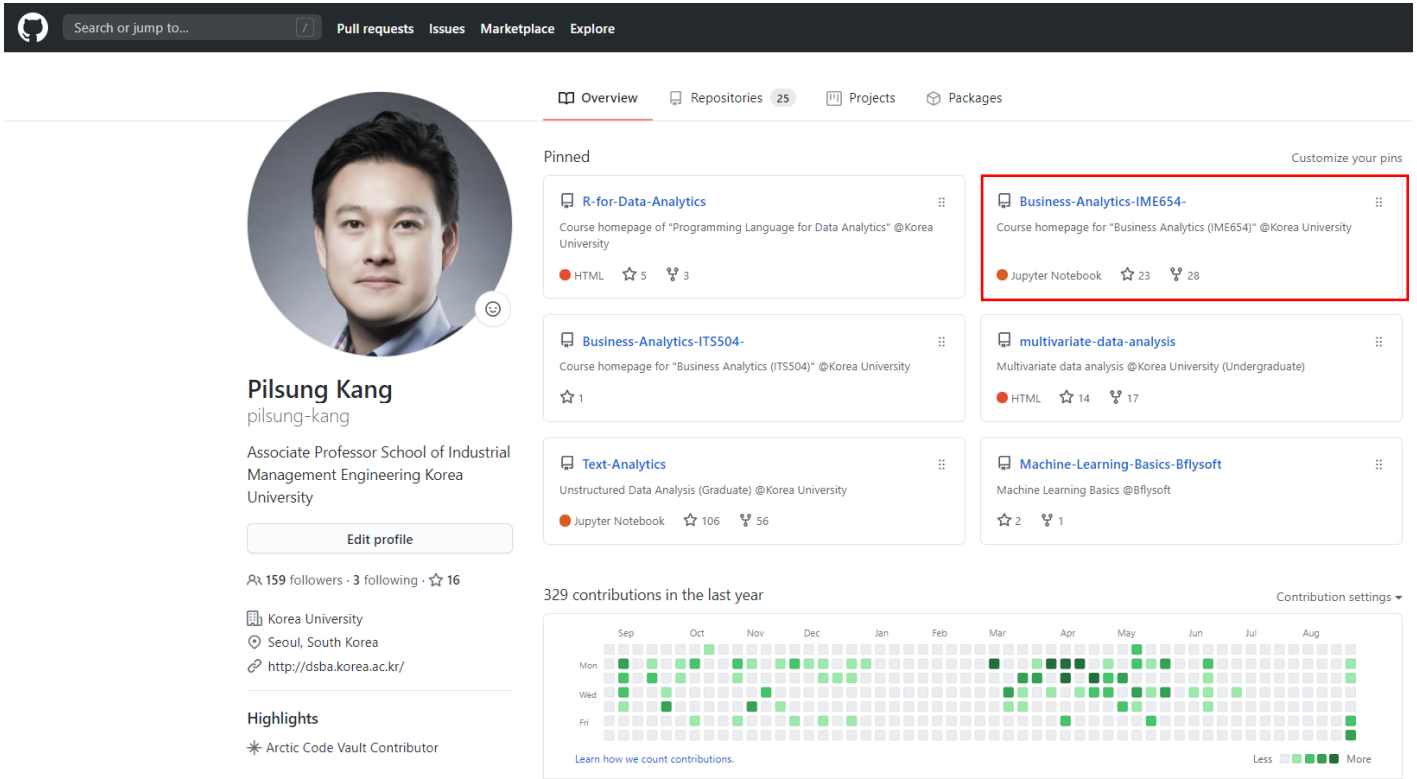
- Course Description
  - ✓ This course provides advanced topics in statistical/machine learning fields
  - ✓ This course also aims at fostering graduate students to have Python programming skills to implement the introduced (or related) algorithms during the lecture

Overview

- Topics
  - ✓ Dimensionality reduction: forward/backward/stepwise selection, genetic algorithm, principal component analysis (PCA), multidimensional scaling (MDS), locally linear embedding, ISOMAP, t-SNE, etc.
  - ✓ Kernel-based learning: support vector machine (SVM), support vector regression (SVR), Kernel Fisher discriminant analysis (KFDA), Kernel principal component analysis (KPCA), etc.
  - ✓ Novelty detection: Gaussian density estimation, mixture of Gaussians, Parzen window density estimation, 1-SVM, SVDD, local outlier factor (LOF), iForests, Robust Cut Forest, etc.
  - ✓ Ensemble learning: Bagging, AdaBoost, Gradient Boosting Machine (GBM), XGBoost, CatBoost, Random Forests, etc.
  - ✓ Semi-supervised learning: Self-training, Generative models, semi-supervised SVM, Graph-based SSL, Co-training, (Re)Mixmatch, Fixmatch, etc.

Lecturer & Course Homepage

- Pilsung Kang, Associate professor at School of Industrial Management Engineering, Korea University
  - ✓ E-mail: pilsung_kang@korea.ac.kr
  - ✓ Course homepage: https://github.com/pilsung-kang/Business-Analytics-IME654-



Lecturer & Course Homepage

master · 1 branch · 0 tags · Go to file · Add file · Code

pilsung-kang Update README.md a1fa97d 2 minutes ago 150 commits

01 Dimensionality Reduction	update materials	4 minutes ago
Lecture Notes (2018)	Update materials	12 months ago
Lecture Notes (2019)	update materials	4 minutes ago
README.md	Update README.md	2 minutes ago

README.md

Business-Analytics

Course homepage for "Business Analytics" @Korea University

Notice

- Syllabus (Slide, Video)
- Tutorial resources (2015)
- Tutorial resources (2016)

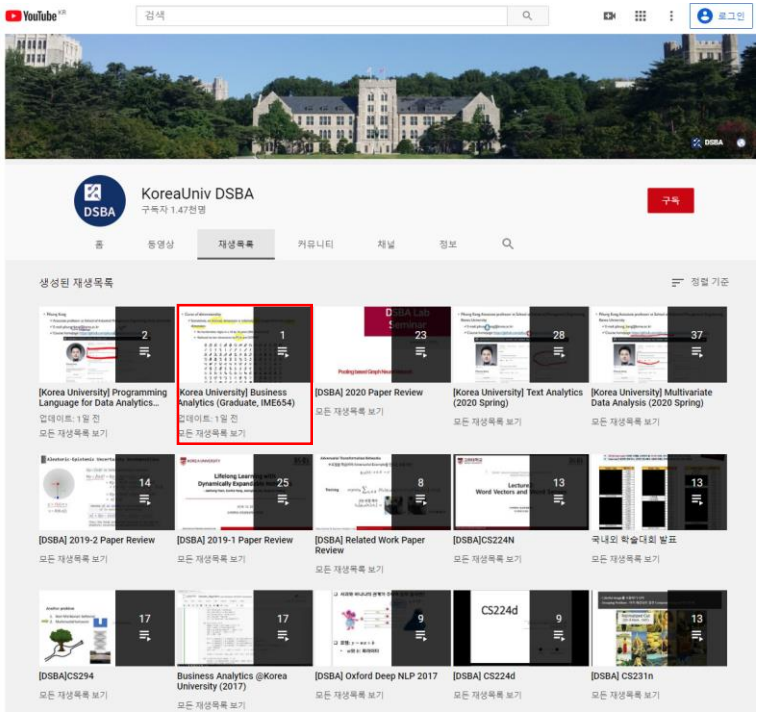
Schedule

Topic 1: Dimensionality Reduction

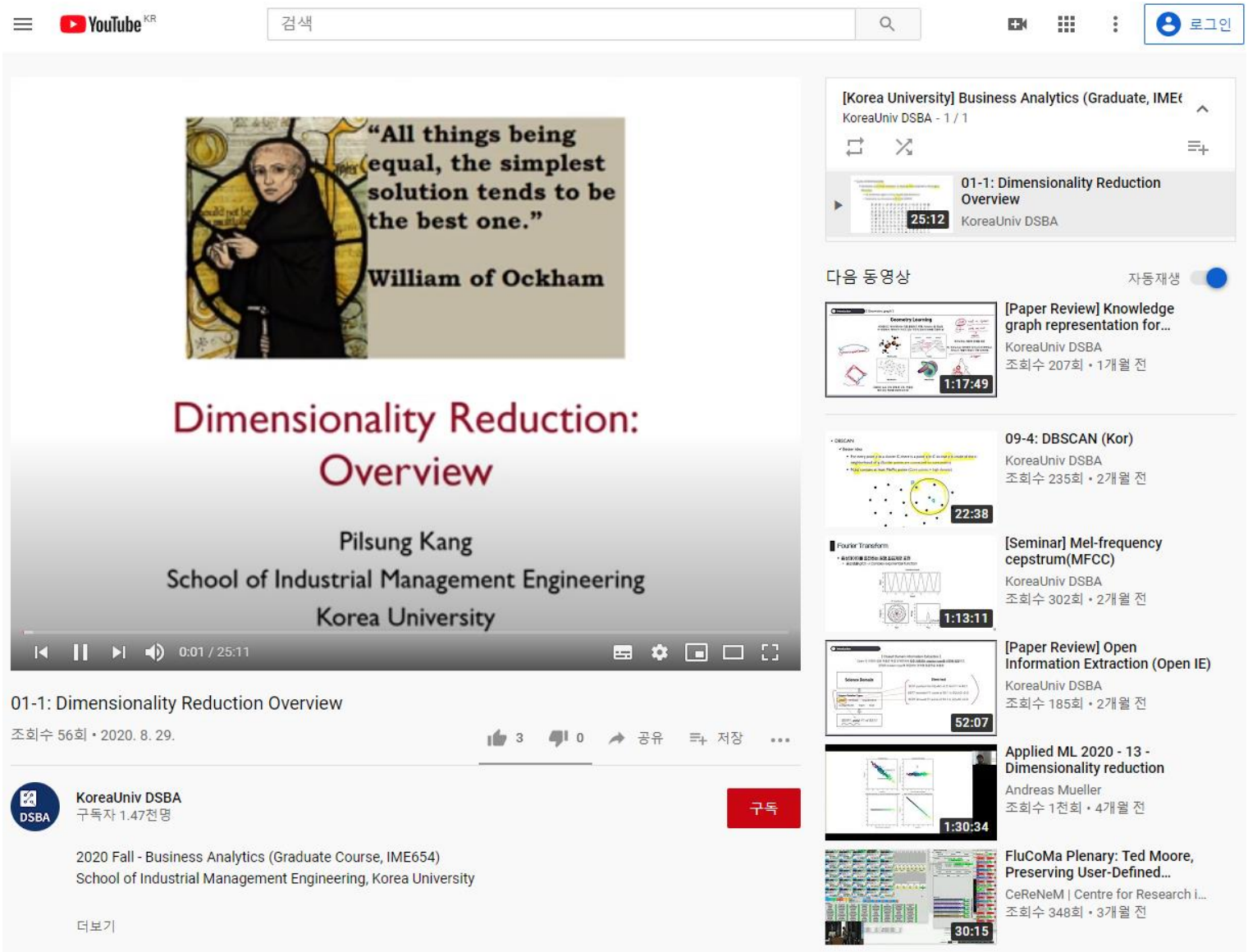
- Dimensionality Reduction: Overview
- Supervised Methods: Forward selection, Backward elimination, Stepwise selection, Genetic algorithm
- Unsupervised Method (Linear embedding): Principal component analysis (PCA), Multi-dimensional scaling (MDS)
- Unsupervised Method (Nonlinear embedding): ISOMAP, LLE, t-SNE
- Tutorial 1: Supervised Method
  - Tutorial Video (장명준), Tutorial Notes (최희정), Tutorial Notes (박경찬)
- Tutorial 2: Unsupervised Method (Linear embedding)
  - Tutorial Video (정재윤), Tutorial Notes (천우진), Tutorial Notes (이정호)
- Tutorial 3: Unsupervised Method (Nonlinear embedding)
  - Tutorial Video (서승완), Tutorial Notes (양우식), Tutorial Notes (손규빈)

Lecture Video

- DSBA Lab Youtube Channel
  - ✓ https://www.youtube.com/channel/UCPq01cgCcEwhXl7BvcwIQyg
- Playlist for this course
  - ✓ https://www.youtube.com/channel/UCPq01cgCcEwhXl7BvcwIQyg/playlists?view_as=subscriber

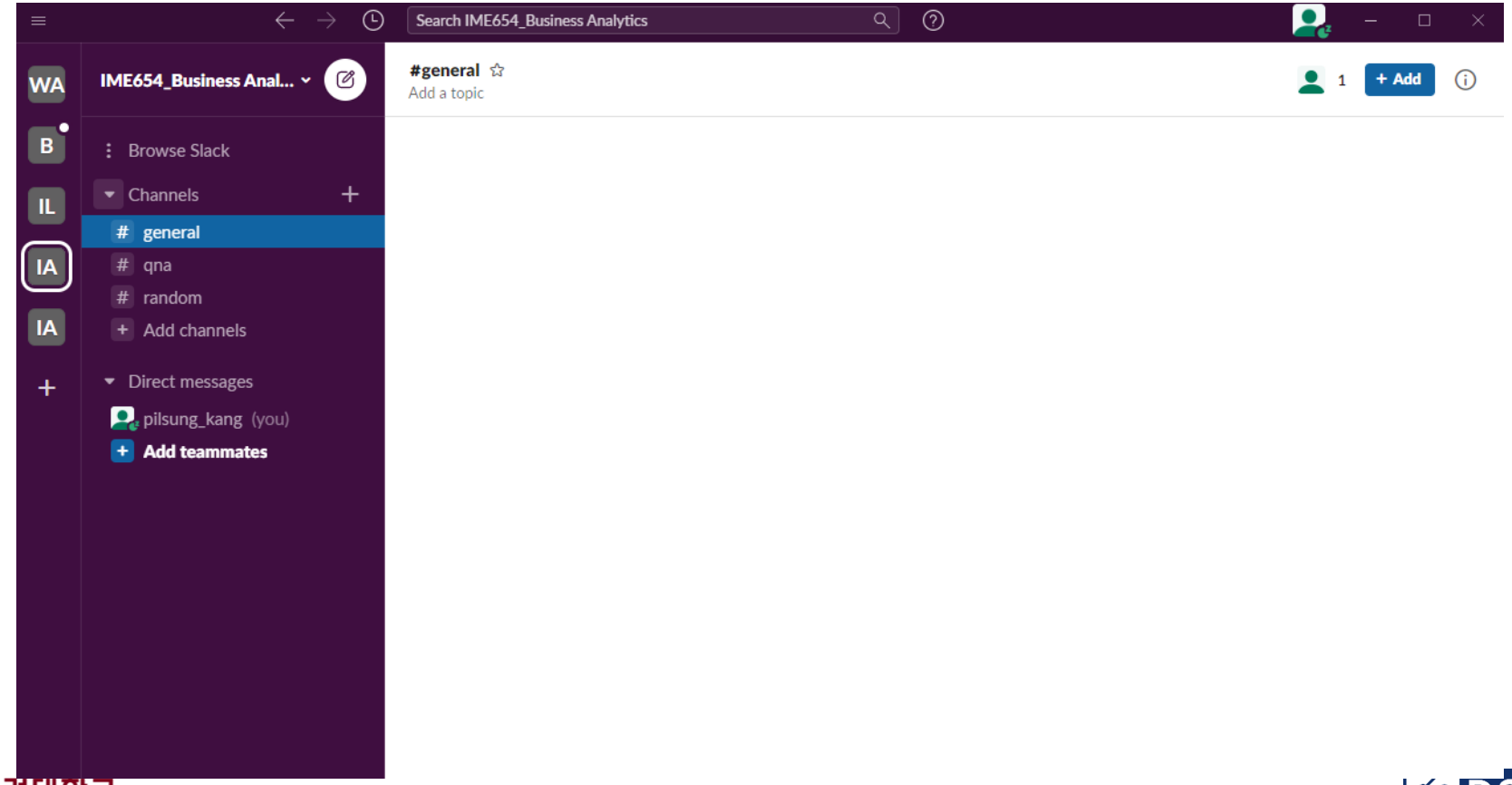


Lecture Video



Communication Channel

- Slack will be used for real-time communication channel
  - ✓ ime654-koreauniv.slack.com
  - ✓ The invitation link will be sent to the enrolled students via e-mail



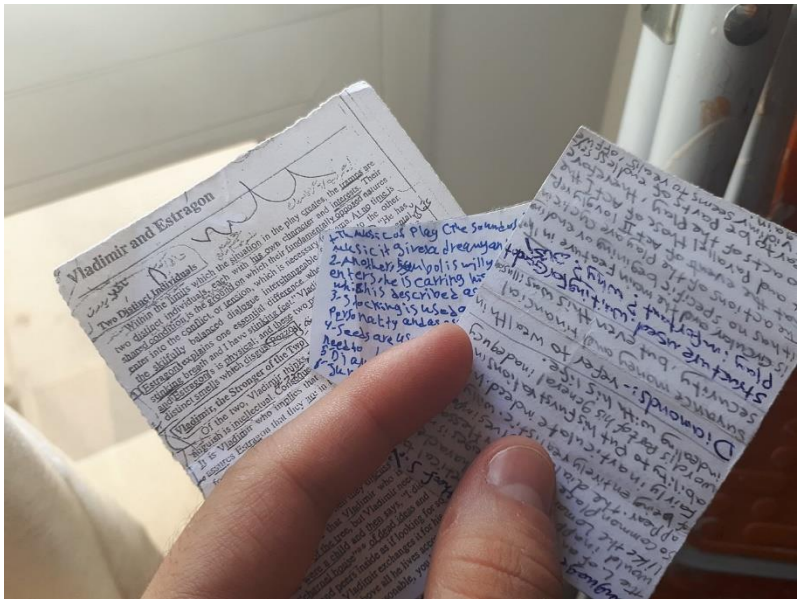
Lecture Modules & Self-Introduction

- Textbook
  - ✓ No textbook is needed.
  - ✓ Lecture notes (PDF format) and recommended paper lists will be provided.

- Introduce Yourself
  - ✓ Submit your self-introduction slide (max. 5 pages) to the lecturer via E-mail (due date: 2020-09-11)
  - ✓ Required information: Name, department, e-mail, cell phone number, recent photo(s)

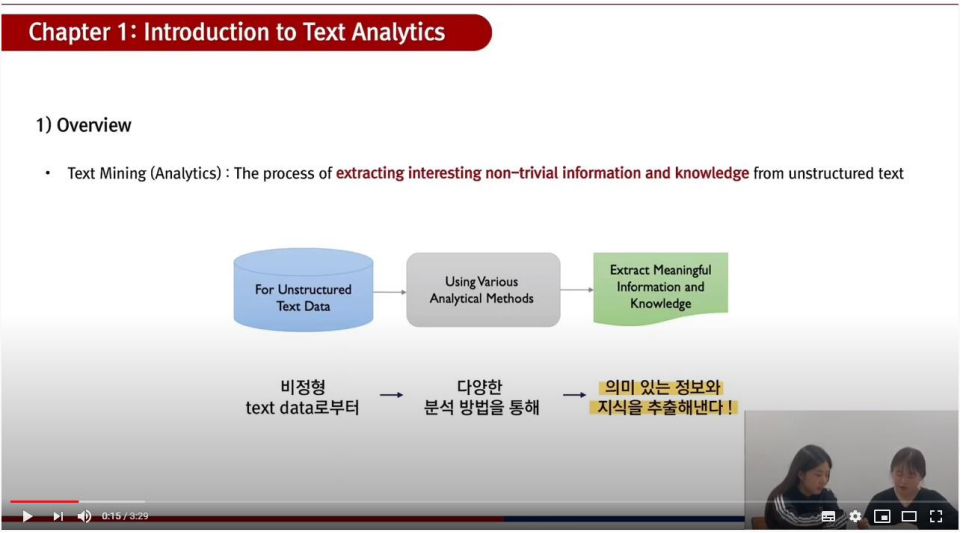
Assessments

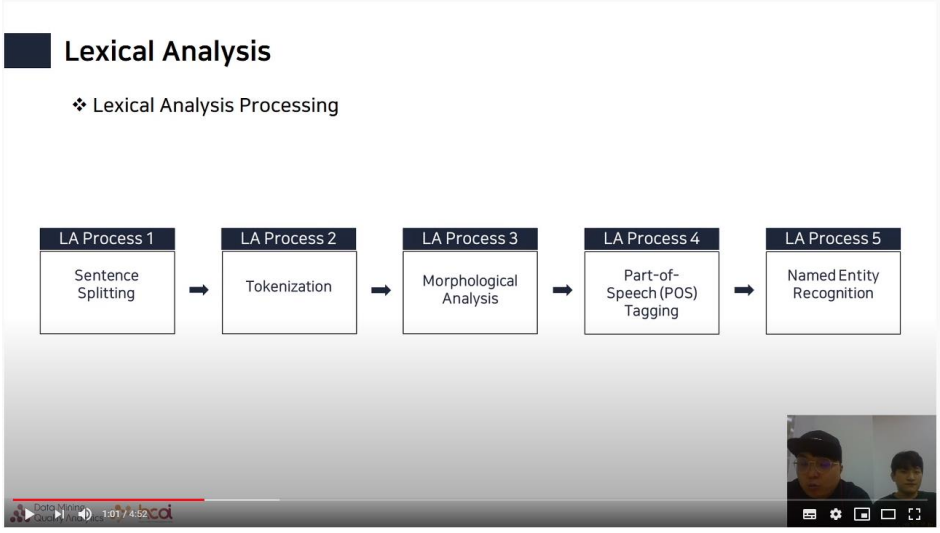
- Final exam (25%)
  - ✓ Three pages of cheating sheets are allowed.



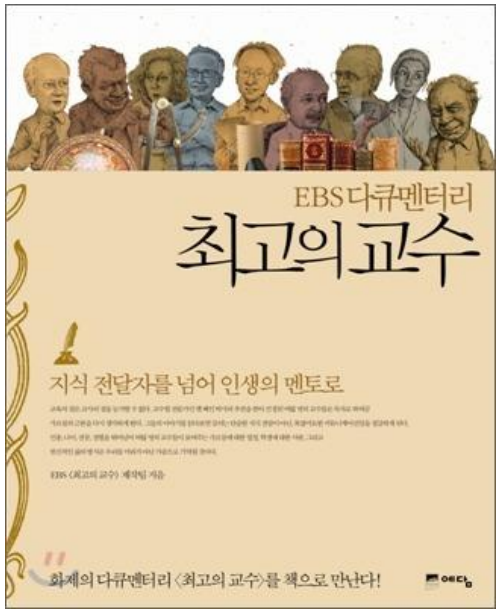
Assessments

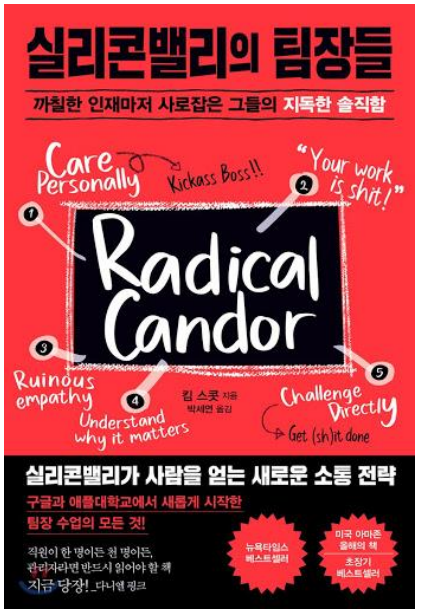
- 5-minutes Youtube video (25%)
  - ✓ Students must upload a short video (max 5 minutes) that reviews the topic covered in the scheduled lecture within 48 hours after the class.
  - ✓ A student must explain what he/she learns in the class to his/her partner.





어떤 교수가 될 것인가?





EBS 다큐프라임 5부작: 최고의 교수

어떤 교수가 될 것인가?

왜 우리는 대학에 가는가

대한민국 청춘들에게 던지는 질문

제18회 YWVCA가 뽑은 좋은 TV프로그램상 특별상(한국YWCA연합회)

[총 5개의 VOD]

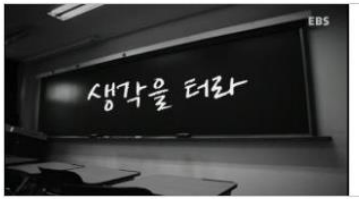
~~5,000~~ (67% 할인) ₩ 1,670

포인트 결제 EBS머니 결제

· 구매일로부터 7일까지 사용 가능합니다.
· 부분환불이 불가한 상품입니다.

결제취소/환불안내 · 이용약관

장바구니 구매하기



왜 우리는 대학에 가는가 - 6부 생각을 터라

다큐프라임

왜 우리는 대학에 가는가 - 4부 어메이징 데이 II

다큐프라임

왜 우리는 대학에 가는가 - 3부 인재의 탄생

다큐프라임

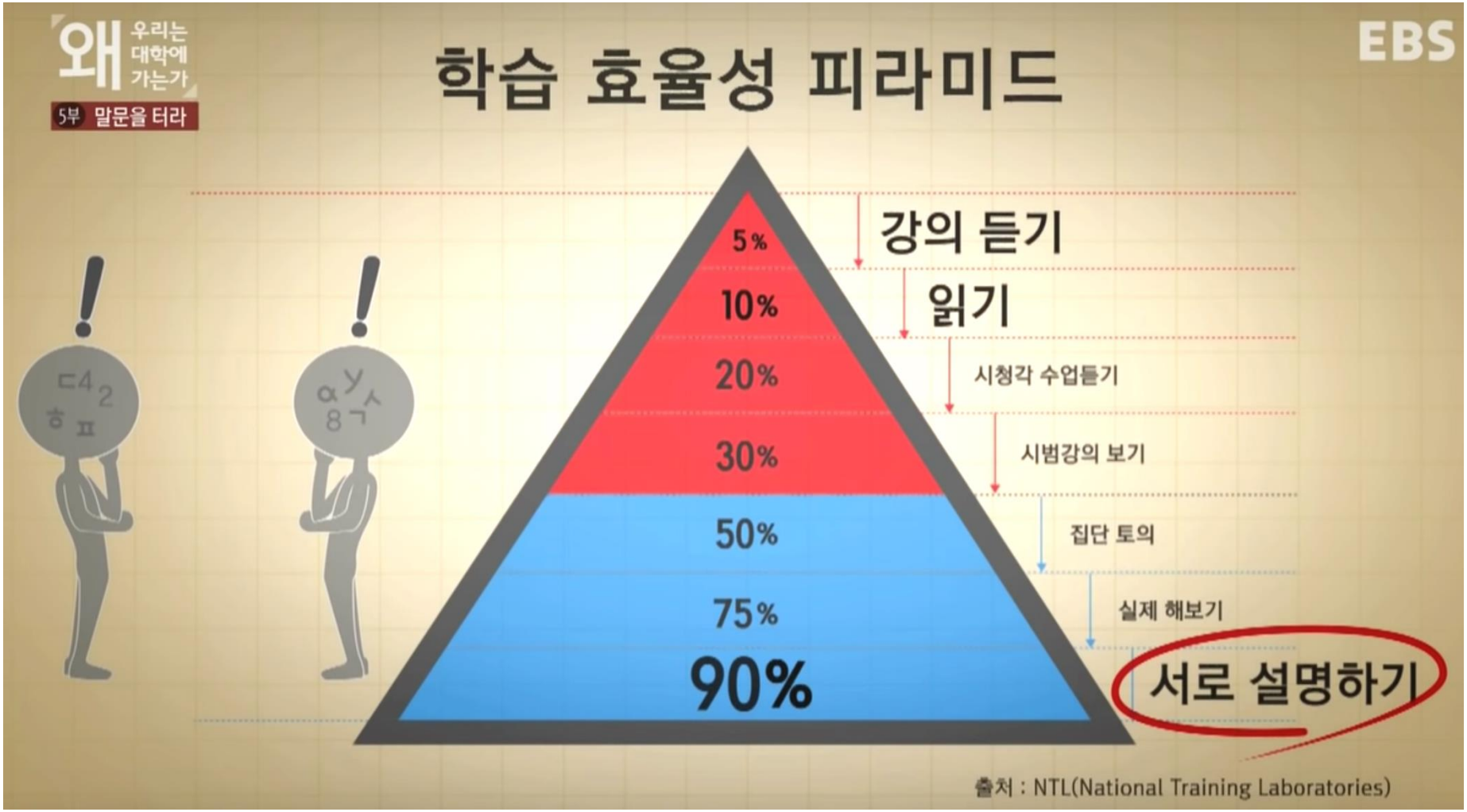
왜 우리는 대학에 가는가 - 2부 인재의 탄생

다큐프라임

왜 우리는 대학에 가는가 - 1부 어메이징 데이

다큐프라임

어떤 교수가 될 것인가?


왜 우리는 대학에 가는가
5부 말문을 터라
EBS
학습 효율성 피라미드
5%
강의 듣기
10%
읽기
20%
시청각 수업들기
30%
시범강의 보기
50%
집단 토의
75%
실제 해보기
90%
서로 설명하기
출처 : NTL(National Training Laboratories)

Assessments

- 5-minute Youtube videos (25%)
  - ✓ Students must upload a short video (max 5 minutes) that reviews the topic covered in the scheduled lecture within 48 hours after the class.
  - ✓ A student must explain what he/she learns in the class to his/her partner.

9) 교수님이 강의 준비를 꼼꼼히 하신다는게 느껴져서 좋았습니다
10) 체계적인 수업 내용, 어디서도 자세하게 알 수 없는 내용 들을 알 수 있어서 좋았습니다.
11) 이론적으로 상세히 알려주시고 유튜브 과제를 통해 학습하는데 도움을 많이 얻었던거 같습니다.
12) 세세하게 A부터 Z까지 설명해주시는 교수님의 강의방식과 레퍼런스, 추가적인 공부자료를 남겨주셔서 많은 도움이 됐습니
13) 교수님께서 친절하게 잘 설명해주셨습니다
14) .
15) 자세한 설명방법과 온라인으로 진행되어도 너무 얻을 것이 많은 과목이었습니다.
16)
17) 수업내용 요약해서 영상찍어 올리는 부분이 힘들었지만 새로운 경험이었습니다.
18) 매 강의 마다 동료에게 수업하고, 유튜브에 업로드한 방식 솔직히 매우 귀찮았지만, 지식 습득에 도움이 된다는 것은 인정할 수 밖에 없다.
19) 한 학기동안 좋은 강의 감사드립니다.
20) 강의가 너무 좋았습니다.
21) 수준높은 강의
22) .
23) 텍스트마이닝과 자연어처리에 대하여 꼼꼼하게 가르쳐주셔서 좋았습니다.

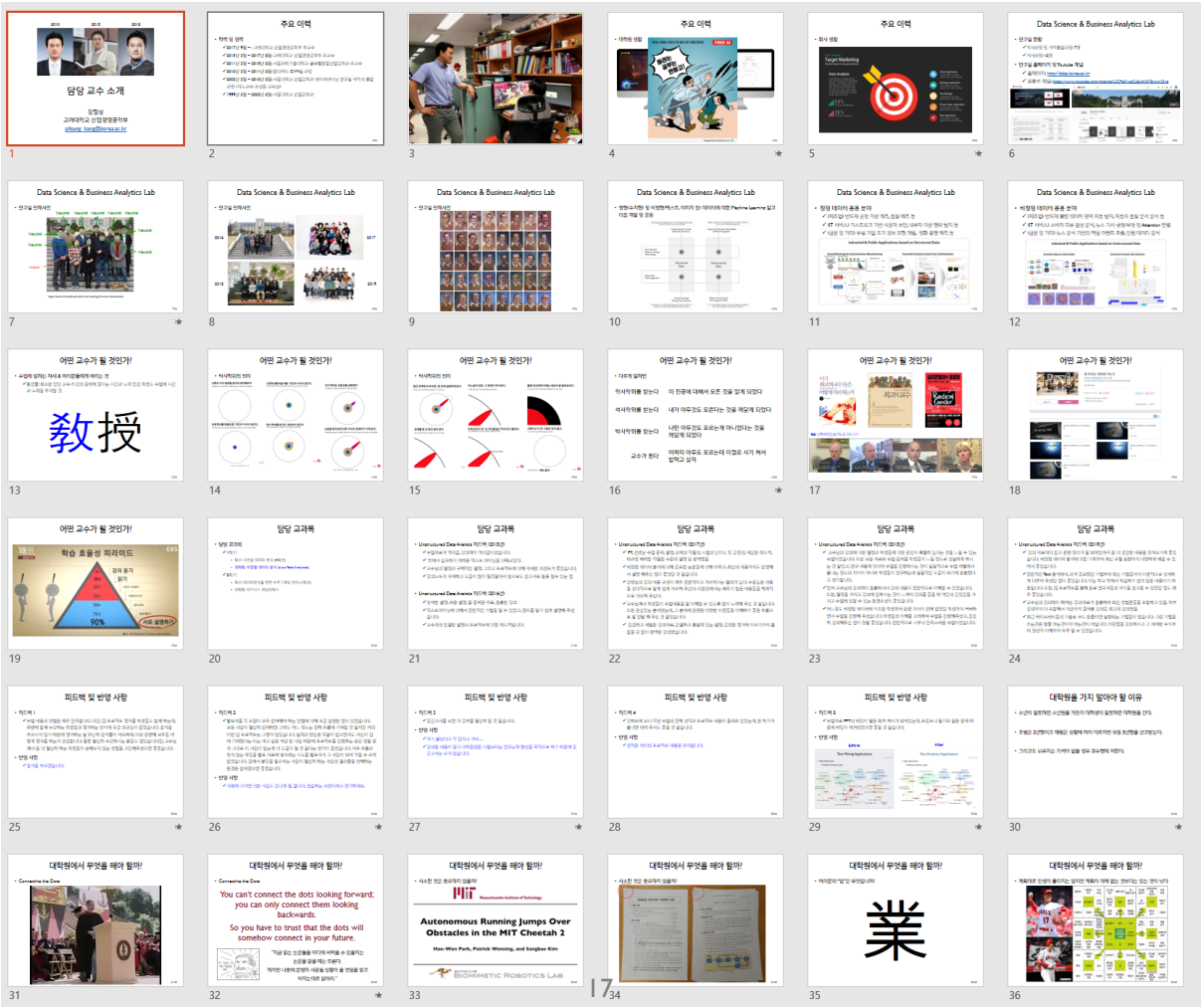
Assessments

- Research paper reproduction (50%)
  - ✓ Students are required to reproduce one excellent (at least good) paper with Python for each topic category (5 reproductions in total).
    - How to determine how good this paper is?
      - https://www.notion.so/9568e19aed0641c09cd02150b3cb3114
  - ✓ The contents of the selected paper and corresponding python code should be explained in a single web page post.
    - Example 1: https://machinelearningmastery.com/machine-learning-in-python-step-by-step/
    - Example 2: https://www.analyticsvidhya.com/blog/2019/11/comprehensive-guide-attention-mechanism-deep-learning/

Introduction to Yourself

✓ Submit your self-introduction slide (max. 5 pages) to the lecturer via E-mail

- Due date: 2020-09-11



Schedule

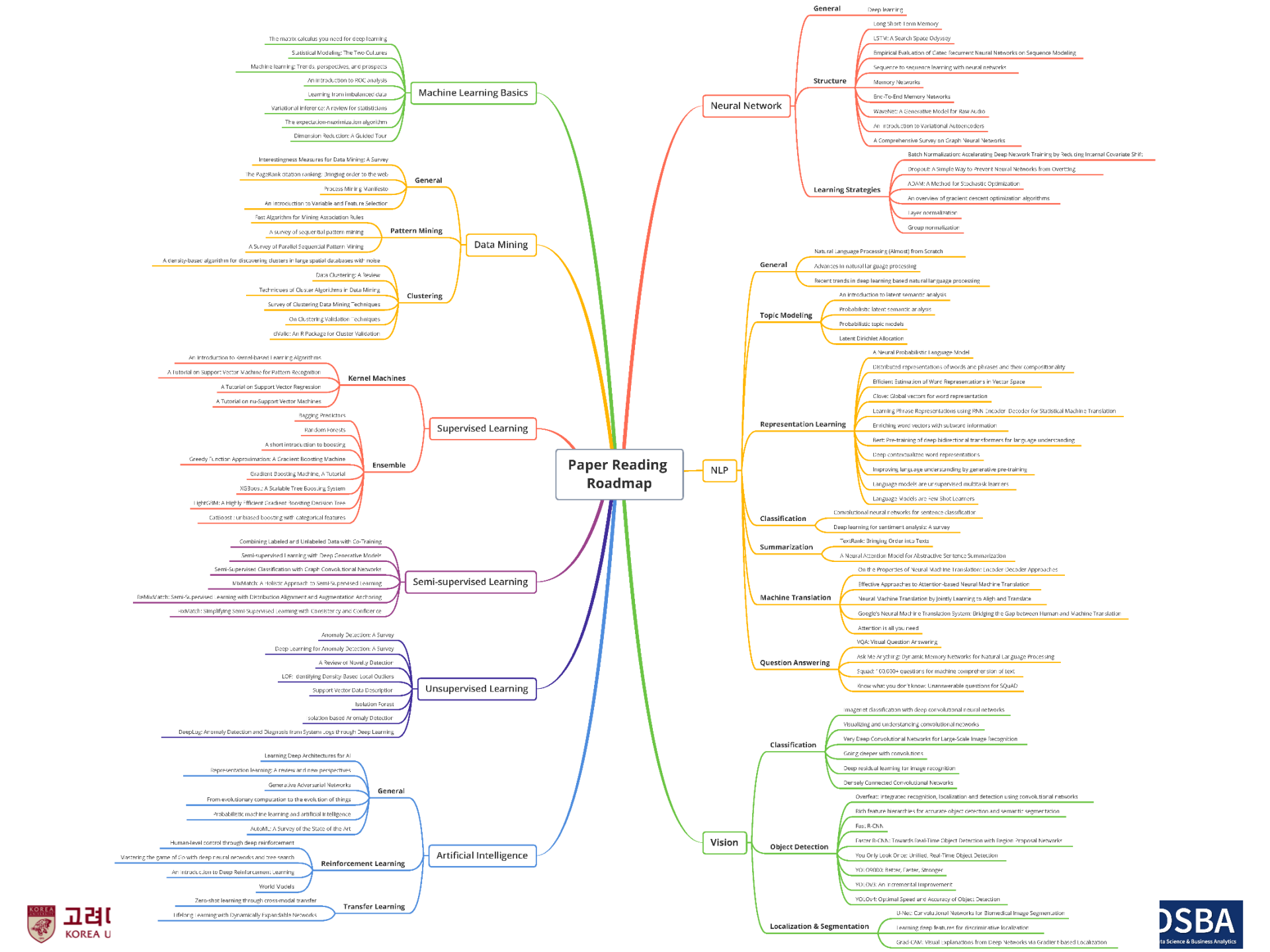
Week	Topics
1	Orientation
2	Dimensionality reduction: forward/backward/stepwise selection, genetic algorithm, PCA
3	Dimensionality reduction: MDS, ISOMAP, LLE, t-SNE
4	Kernel-based learning: SVM
5	Kernel-based learning: SVR, KFDA, KPCA
6	Novelty detection: Gauss, MoG, Parzen
7	Novelty detection: k-NN, LOF, 1-SVM, SVDD
8	No Class
9	Novelty detection: PCA-based, Clustering-based, iForest, Robust Random Cut Forest
10	Ensemble learning: Bagging, Random Forests
11	Ensemble learning: AdaBoost, GBM
12	Ensemble learning: XGBoost, CatBoost
13	Semi-supervised learning: Self-training, Generative models
14	Semi-supervised learning: SS-SVM, Graph-based SSL
15	Semi-supervised learning: Co-Training, (Re)MixMatch, FixMatch
16	Final Exam

Additional Material

- If you are the first semester in the Graduate school
  - ✓ https://www.notion.so/c3b3474d18ef4304b23ea360367a5137?v=5d763ad5773f44eb950f49de7d7671bd

Papers You Must Read

Title	Journal/Conference	Year	Description	태그	파일	XMind
Language Models are Few-Shot Learners	arXiv	2020	GPT-3	NLP Neural Network	2005.14165.pdf	Included
FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence	arXiv	2020	FixMatch	Semi-supervised Learning	2001.07685.pdf	Included
YOLOv4: Optimal Speed and Accuracy of Object Detection	arXiv	2020	Yolo-V4	Deep Learning Vision	2004.10934.pdf	Included
A Comprehensive Survey on Graph Neural Networks	IEEE TNNLS	2020	GNN	Graph Neural Network	1901.00596.pdf	Included
Language models are unsupervised multitask learners		2019	GPT-2	NLP Neural Network	2019-Radford-...	Included
MixMatch: A Holistic Approach to Semi-Supervised Learning	NIPS	2019	MixMatch	Semi-supervised Learning	1905.02249.pdf	Included
ReMixMatch: Semi-Supervised Learning with Distribution Alignment and Augmentation Anchoring	arXiv	2019	ReMixMatch	Semi-supervised Learning	1911.09785.pdf	Included
AutoML: A Survey of the State-of-the-Art	arXiv	2019	Auto-ML	Auto-ML	AutoML_A Surv...	Included
Deep Learning for Anomaly Detection: A Survey	arXiv	2019	Anomaly Detection	Deep Learning Anomaly Detection Survey	1901.03407.pdf	Included
An Introduction to Variational Autoencoders	FTML	2019	Variational Autoencoder	Neural Network	1906.02691.pdf	Included
A Survey of Parallel Sequential Pattern Mining	ACM TKDD	2019	Sequential Pattern Mining	Data Mining	3314107.pdf	Included
Improving language understanding by generative pre-training		2018	GPT	NLP Neural Network	language_unde...	Included
Deep contextualized word representations	arXiv	2018	ELMo	NLP Neural Network	1802.05365.pdf	Included
Bert: Pre-training of deep bidirectional transformers for language understanding	arXiv	2018	BERT	NLP Neural Network	1810.04805.pdf	Included
Deep learning for sentiment analysis: A survey		2018	Sentiment Analysis	NLP Survey	1801.07883.pdf	Included
CatBoost : unbiased boosting with categorical features	NIPS	2018	CatBoost	Ensemble Learning	CatBoost_unbia...	Included
The matrix calculus you need for deep learning		2018	Matrix Calculus	Machine Learning Basics	The matrix calc...	Included
YOLOv3: An Incremental Improvement	arXiv	2018	Yolo-V3	Deep Learning Vision	1804.02767.pdf	Included
Know what you don't know: Unanswerable questions for SQuAD	arXiv	2018	Question Answering	NLP	1806.03822.pdf	Included
Group normalization	ECCV	2018	Normalization	Neural Network	Yuxin_Wu_Grou...	Included


Paper Reading Roadmap
Machine Learning Basics
The matrix calculus you need for deep learning
Statistical Modeling: The Two Cultures
Machine learning: Trends, perspectives, and prospects
An introduction to ROC analysis
Learning from imbalanced data
Variational Inference: A review for statisticians
The expectation-maximization algorithm
Dimension Reduction: A Guided Tour
Data Mining
General
Interestingness Measures for Data Mining: A Survey
The PageRank citation ranking: Bringing order to the web
Process Mining Manifesto
An Introduction to Variable and Feature Selection
Pattern Mining
Fast Algorithm for Mining Association Rules
A survey of sequential pattern mining
A Survey of Parallel Sequential Pattern Mining
Clustering
A density-based algorithm for discovering clusters in large spatial databases with noise
Data Clustering: A Review
Techniques of Cluster Algorithms in Data Mining
Survey of Clustering Data Mining Techniques
On Clustering Validation Techniques
dVallic: An R Package for Cluster Validation
Supervised Learning
Kernel Machines
An Introduction to Kernel-based Learning Algorithms
A Tutorial on Support Vector Machine for Pattern Recognition
A Tutorial on Support Vector Regression
A Tutorial on nu-Support Vector Machines
Ensemble
Bagging Predictors
Random Forests
A short introduction to boosting
Greedy Function Approximation: A Gradient Boosting Machine
Gradient Boosting Machine, A Tutorial
XGBoost: A Scalable Tree Boosting System
LightGBM: A Highly Efficient Gradient Boosting Decision Tree
CatBoost: unbiased boosting with categorical features
Semi-supervised Learning
Combining Labeled and Unlabeled Data with Co-Training
Semi-supervised Learning with Deep Generative Models
Semi-Supervised Classification with Graph Convolutional Networks
MixMatch: A Holistic Approach to Semi-Supervised Learning
ReMixMatch: Semi-Supervised Learning with Distribution Alignment and Augmentation Anchoring
FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence
Unsupervised Learning
Anomaly Detection: A Survey
Deep Learning for Anomaly Detection: A Survey
A Review of Novelty Detection
LOF: Identifying Density-Based Local Outliers
Support Vector Data Description
Isolation Forest
Isolation-based Anomaly Detection
DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning
Artificial Intelligence
General
Learning Deep Architectures for AI
Representation learning: A review and new perspectives
Generative Adversarial Networks
From evolutionary computation to the evolution of things
Probabilistic machine learning and artificial intelligence
AutoML: A Survey of the State-of-the-Art
Reinforcement Learning
Human-level control through deep reinforcement
Mastering the game of Go with deep neural networks and tree search
An Introduction to Deep Reinforcement Learning
World Models
Transfer Learning
Zero-shot learning through cross-modal transfer
Lifelong Learning with Dynamically Expandable Networks
Neural Network
General
Deep learning
Structure
Long Short-Term Memory
LSTM: A Search Space Odyssey
Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling
Sequence to sequence learning with neural networks
Memory Networks
End-To-End Memory Networks
WaveNet: A Generative Model for Raw Audio
An Introduction to Variational Autoencoders
A Comprehensive Survey on Graph Neural Networks
Learning Strategies
Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift
Dropout: A Simple Way to Prevent Neural Networks from Overfitting
ADAM: A Method for Stochastic Optimization
An overview of gradient descent optimization algorithms
Layer normalization
Group normalization
NLP
General
Natural Language Processing (Almost) from Scratch
Advances in natural language processing
Recent trends in deep learning based natural language processing
Topic Modeling
An Introduction to latent semantic analysis
Probabilistic latent semantic analysis
Probabilistic topic models
Latent Dirichlet Allocation
Representation Learning
A Neural Probabilistic Language Model
Distributed representations of words and phrases and their compositionality
Efficient Estimation of Word Representations in Vector Space
Glove: Global vectors for word representation
Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation
Enriching word vectors with subword information
Bert: Pre-training of deep bidirectional transformers for language understanding
Deep contextualized word representations
Improving language understanding by generative pre-training
Language models are unsupervised multitask learners
Language Models are Few-Shot Learners
Classification
Convolutional neural networks for sentence classification
Deep learning for sentiment analysis: A survey
Summarization
TextRank: Bringing Order into Texts
A Neural Attention Model for Abstractive Sentence Summarization
Machine Translation
On the Properties of Neural Machine Translation: Encoder-Decoder Approaches
Effective Approaches to Attention-based Neural Machine Translation
Neural Machine Translation by Jointly Learning to Align and Translate
Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation
Attention is all you need
Question Answering
VQA: Visual Question Answering
Ask Me Anything: Dynamic Memory Networks for Natural Language Processing
Squad: 100,000+ questions for machine comprehension of text
Know what you don't know: Unanswerable questions for SQuAD
Vision
Classification
Imagenet classification with deep convolutional neural networks
Visualizing and understanding convolutional networks
Very Deep Convolutional Networks for Large-Scale Image Recognition
Going deeper with convolutions
Deep residual learning for image recognition
Densely Connected Convolutional Networks
Object Detection
Overfeat: Integrated recognition, localization and detection using convolutional networks
Rich feature hierarchies for accurate object detection and semantic segmentation
Fast R-CNN
Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks
You Only Look Once: Unified, Real-Time Object Detection
YOLO9000: Better, Faster, Stronger
YOLOv3: An Incremental Improvement
YOLOv4: Optimal Speed and Accuracy of Object Detection
Localization & Segmentation
U-Net: Convolutional Networks for Biomedical Image Segmentation
Learning deep features for discriminative localization
Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization