|  |  |
| --- | --- |
| 교육 제목 | **미분 복습 & Regrssion & ML** |
| 교육 일시 | 21.10.13 |
| 교육 장소 | 오프라인 (영우글로벌러닝) |
| **교육 내용** | |
| 오전 | **▣ Regression**  -------------------------------------------------------------------------------------------  **∇ 최소제곱추정 (OLS)**  \* 에러가 적은 추정량일수록 좋음  \* 에러가 가정 적은 추정량을 구함  -------------------------------------------------------------------------------------------  **∇ Loss Function (Parameter Estimation)**  \* 에러가 음수일 수 있기 때문에 제곱의 합으로 계산함.  \* 미분시 0이 되는 상수 값(오메가)들을 구함  -------------------------------------------------------------------------------------------  **∇ python에서의 활용**  \* 데이터 (feature)들을 설정 정의 해줌.  \* S\_xx와 S\_xy 를 구해 parameter를 estimation 함.  \* y의 추정량을 구함 및 그래프 plot  \* plt.plot() 사용  -------------------------------------------------------------------------------------------  **∇ sklearn 사용**  \* sklearn 중에서 LinearRegression 사용  \* .reshape()으로 matrix 차원 재설정해줌.  \* lr.fit() 으로 두 변수 사이의 관계 설정.  \* lr.coef\_ 는 기울기, lr.intercept\_는 y절편을 구함.  ------------------------------------------------------------------------------------------- |
| 오후 | **▣ Machine Learning**  -------------------------------------------------------------------------------------------  **∇ label 유무에 따른 분류**  \* Supervised Learning: label 포함시.  target이 연속형이냐 범주형이냐에  따라 Regression과 Classification으로 나뉨  \* Unsupervised Learning: label 미포함시.  Clustering과 visualization & Dimension Reduction  그리고 Association Rule Learning으로 나뉨  \* Semisupervised Learning: 위의 두 모델의 조합.  \* Reinforcement Learning: 인간의 학습방법과 가장 유사하다고 볼 수 있음.  -------------------------------------------------------------------------------------------  **∇ Training Set & Test Set**  \* Training Set: 데이터의 70%만을 사용해서 모델 정의  그 중 Validation Set으로 Hyper Parameter, Over-Fit 판단  \* Test Set: 남은 30%를 사용하여 정의된 모델에 대입해서 판단  -------------------------------------------------------------------------------------------  **∇ Bias & Variance**  \* Bias: 편향된 정도를 나타냄. 보통 고려되지 않음  \* Variance: 분산의 정도를 나타냄.  \* Bias와 Variance는 Trade-Off 관계에 있음.  \* 두 변수를 고려해, Training Set에 의해 만들어진 적절한 Error를 찾아내야 함.  -------------------------------------------------------------------------------------------  **∇ Sampling**  \* Sampling: Training Set과 Test Set의 샘플을 만드는 것.  \* Random Sampling: 무작위로 샘플을 만듦  \* Stratified Sampling: 분류 비율을 고려해서 샘플링을 함.  -------------------------------------------------------------------------------------------  **∇ k-fold Cross Validation**  \* k 개의 fold (group)를 형성해서 각각의 training set을 만든 후 교차로 검증  ------------------------------------------------------------------------------------------- |