



# TA Session 1

---

2018.10.12

Wonjong Rhee, Hyunghun Cho, Daeyoung Choi

Seoul National University

Graduate School of Convergence Science and Technology

Applied Data Science Lab.

# Iris data set



50 instances per class

Data Set Characteristics:	Multivariate	Number of Instances:	150	Area:	Life
Attribute Characteristics:	Real	Number of Attributes:	4	Date Donated	1988-07-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	2146288

Attribute Information:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:

- Iris Setosa
- Iris Versicolour
- Iris Virginica



*Iris setosa*



*Iris versicolor*

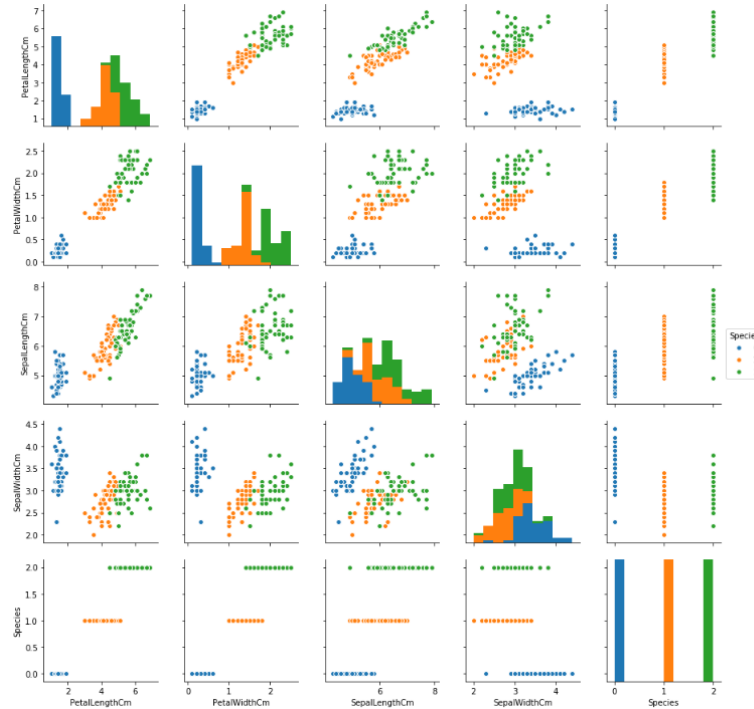


*Iris virginica*

# Task 1: Very simple data analysis



- 아래의 그림을 그려보십시오. (간단한 python function call로 가능) 단, class 정보도 추가하기 바랍니다. 각 independent variable 이 어떤 관계성을 가지는지 서술하세요.



```
import pandas as pd
import seaborn as sns
```

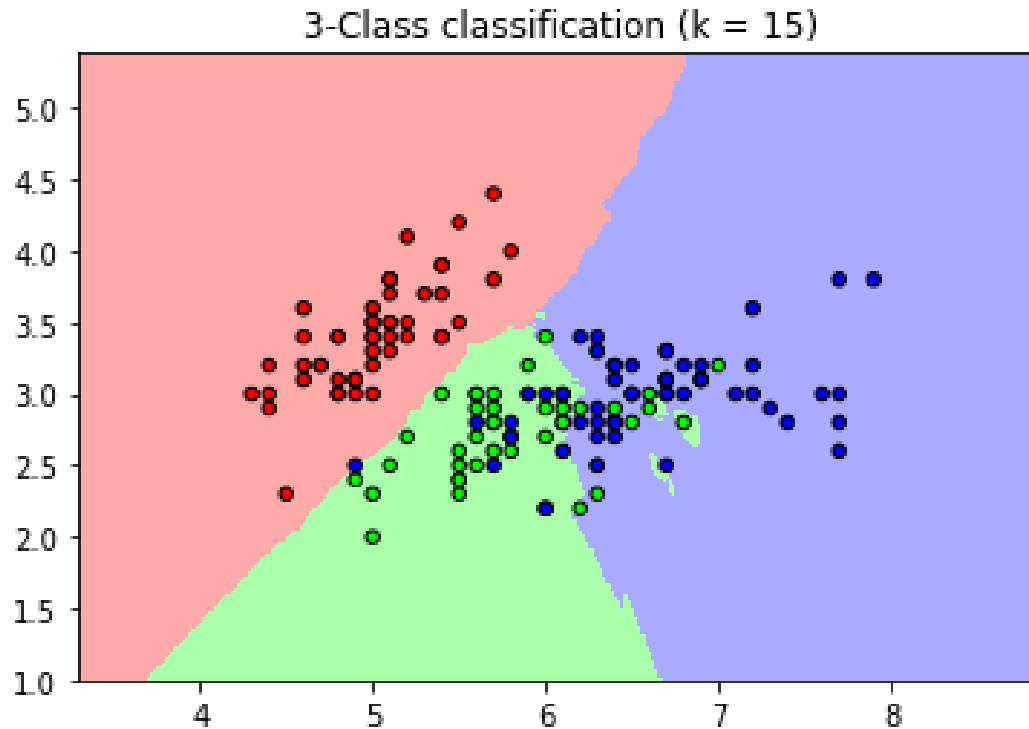
```
df_iris = pd.DataFrame({'SepalLengthCm':iris.data[:, 0], 'SepalWidthCm':iris.data[:, 1], 'PetalLengthCm':iris.data[:, 2], 'PetalWidthCm':iris.data[:, 3], 'Species':iris.target})
print(df_iris)
```

```
sns.pairplot(data=df_iris, hue="Species")
```

## Task 2: Bias-variance tradeoff



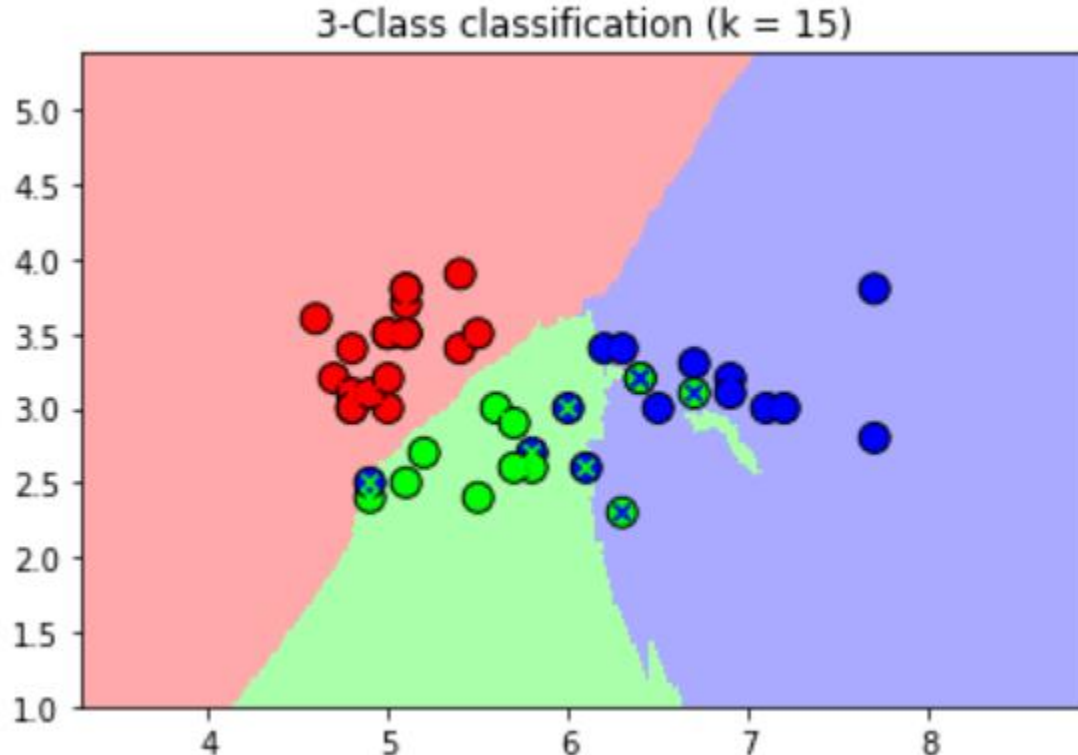
- 주어진 KNN 코드에서  $k$ 를 바꿔가며 아래의 그림을 그려보세요.  $k$ 가 달라짐에 따라 어떤 변화가 있는지 bias-variance tradeoff 관점에서 서술하세요. ( $k$ 는 1,2,5,10,20을 시도하십시오.)



# Task 3: Training-test set split



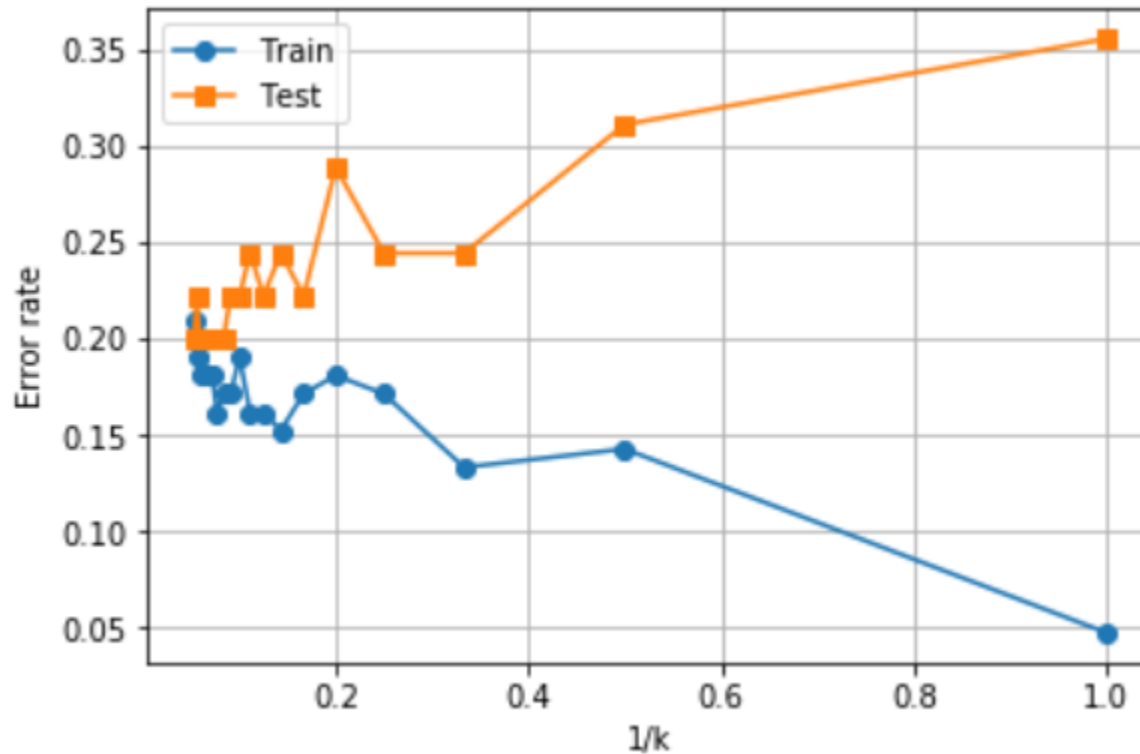
- 주어진 데이터를 training, test set으로 7:3의 비율로 나누고 아래와 같은 그림을 재현해 보십시오. 다음과 같이 하시면 됩니다. 먼저, training set으로 KNN모델을 학습하십시오. 그 다음 test set의 true class label로 Task 2와 같은 그림을 그립니다. 다음으로 test set에 대해 모델이 예측한 class label로 그 위에 'x'로 표현하여 그림을 그립니다. (k는 15로 고정하십시오.)



# Task 4: Overfitting



- Task 3에서 training, test set 으로 나눈 것을 활용하여 아래와 같이  $k$ 가 변함에 따라 training, test error가 어떻게 변하는지 확인할 수 있는 그림을 그려보세요. 결과를 overfitting 관점에서 서술하시오.



- Task는 구현 자체보다는 **결과에 대한 분석과 이해를 중심으로 학습** 하시기 바랍니다. 즉, bias-variance tradeoff, training-test set split, overfitting 등 수업에서 배운 개념을 코드를 짜고 분석하면서 좀 더 이해해 보시기 바랍니다.
- 제공된 2018\_fall\_session\_1\_task.ipynb 파일에 코드를 작성하고 설명을 단 뒤 ipynb 파일을 제출해주세요.
- 형식과 분량은 자유입니다. 제출시간은 진행상황을 보가며 정하도록 하겠습니다.
- 다음의 이메일로 [choid@snu.ac.kr](mailto:choid@snu.ac.kr) (최대영 조교) 로 제출하시고 보내실 때 반드시 **조별 대표자 성명을 기록**해주시기 바랍니다.