딥러닝 - Lab2

명지대학교
MYONGJI UNIVERSITY

과제 제출 유의사항

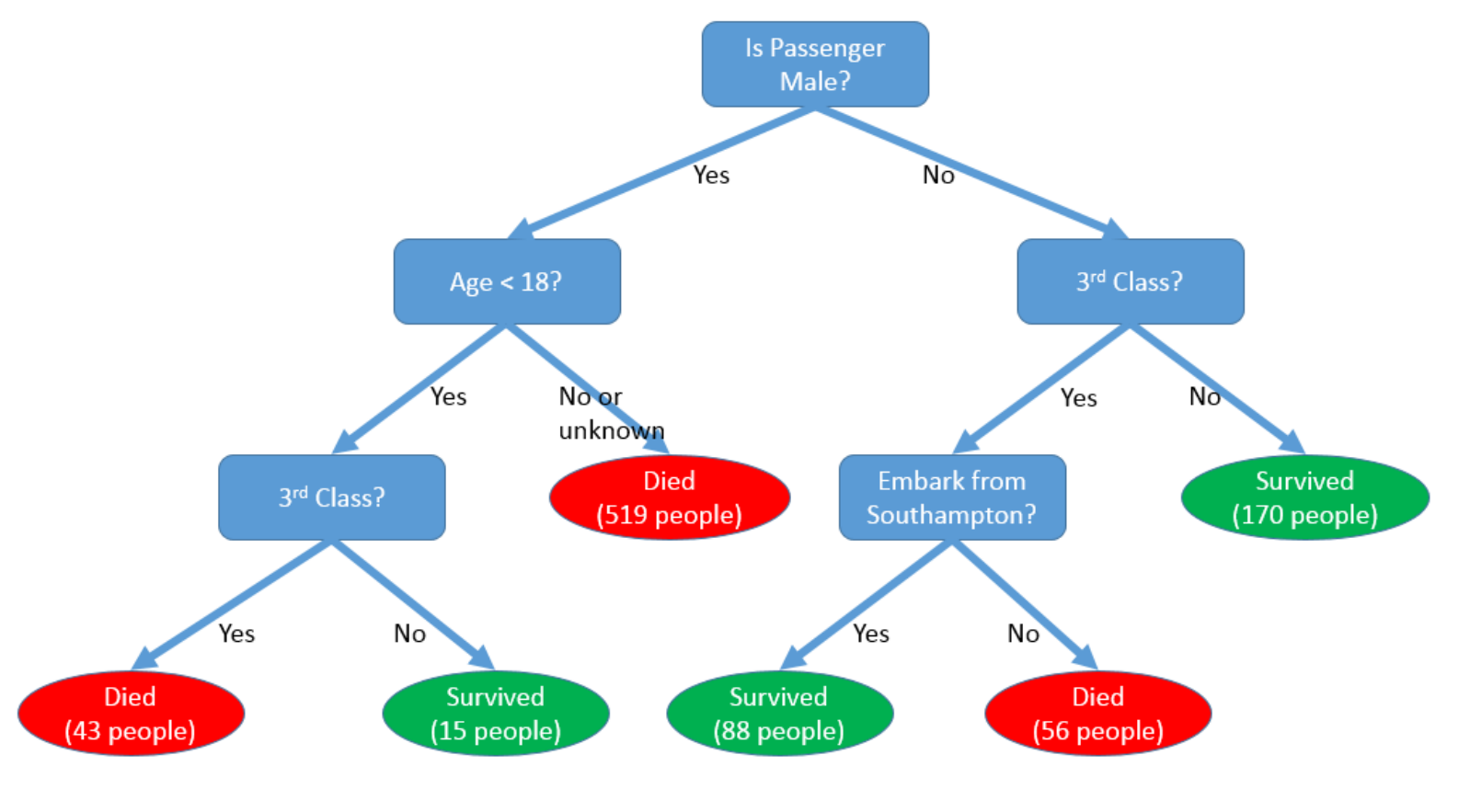
- 이번주 과제는 .ipynb 로 제출 바랍니다.
- 파일에 모든 과제가 다 들어갈 수 있도록 한 개 파일로 작성 (comment 로 구분 필요)
- 과제는 스스로 하기 바랍니다.

Lab2 과제

- **Iris 데이터를 이용하여 머신러닝 방법으로 예측한 것을 비교하여라**
- **단 같은 random seed와 같은 테스트 데이터 셋으로 naïve bayes, decision tree, adaboost을 비교하여라**
  - Naïve bayes (from sklearn.naive_bayes import GaussianNB)ㅍ
    - GaussianNB implements the Gaussian Naive Bayes algorithm for classification. The likelihood of the features is assumed to be Gaussian:Decision tree from sklearn.tree import DecisionTreeClassifier)
  - Decision tree from sklearn.tree import DecisionTreeClassifier)
  - AdaBoost (from sklearn.ensemble import AdaBoostClassifier)
    - n_estimators : 부스팅 종료를 위한 위한 맥시멈 추정값

```
from sklearn import datasets
iris = datasets.load_iris()
```

Decision Tree



https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html#sklearn.tree.DecisionTreeClassifier

Naïve Bayes

- **Bayesian classifiers use Bayes theorem, which says**
- $p(C_j|d) = \frac{p(d|C_j)p(C_j)}{p(d)}$
- **$p(c_j \mid d)$ = probability of instance d being in class c_j,**
- **$p(d \mid c_j)$ = probability of generating instance d given class c_j**
- **$p(c_j)$ = probability of occurrence of class c_j**
- **p(d) = probability of instance d occurring**

https://scikit-learn.org/stable/modules/naive_bayes.html

Adaboost (Adaptive Boosting)

- algorithm that forms a committee of weak classifiers

Algorithm 10.1 *AdaBoost.M1.*

1. Initialize the observation weights $w_i = 1/N,\ i = 1, 2, \ldots, N$.
2. For $m = 1$ to M:
   (a) Fit a classifier $G_m(x)$ to the training data using weights w_i.
   (b) Compute
   $$\text{err}_m = \frac{\sum_{i=1}^{N} w_i I(y_i \neq G_m(x_i))}{\sum_{i=1}^{N} w_i}.$$
   (c) Compute $\alpha_m = \log((1 - \text{err}_m)/\text{err}_m)$.
   (d) Set $w_i \leftarrow w_i \cdot \exp[\alpha_m \cdot I(y_i \neq G_m(x_i))],\ i = 1, 2, \ldots, N$.
3. Output $G(x) = \text{sign}\left[\sum_{m=1}^{M} \alpha_m G_m(x)\right]$.