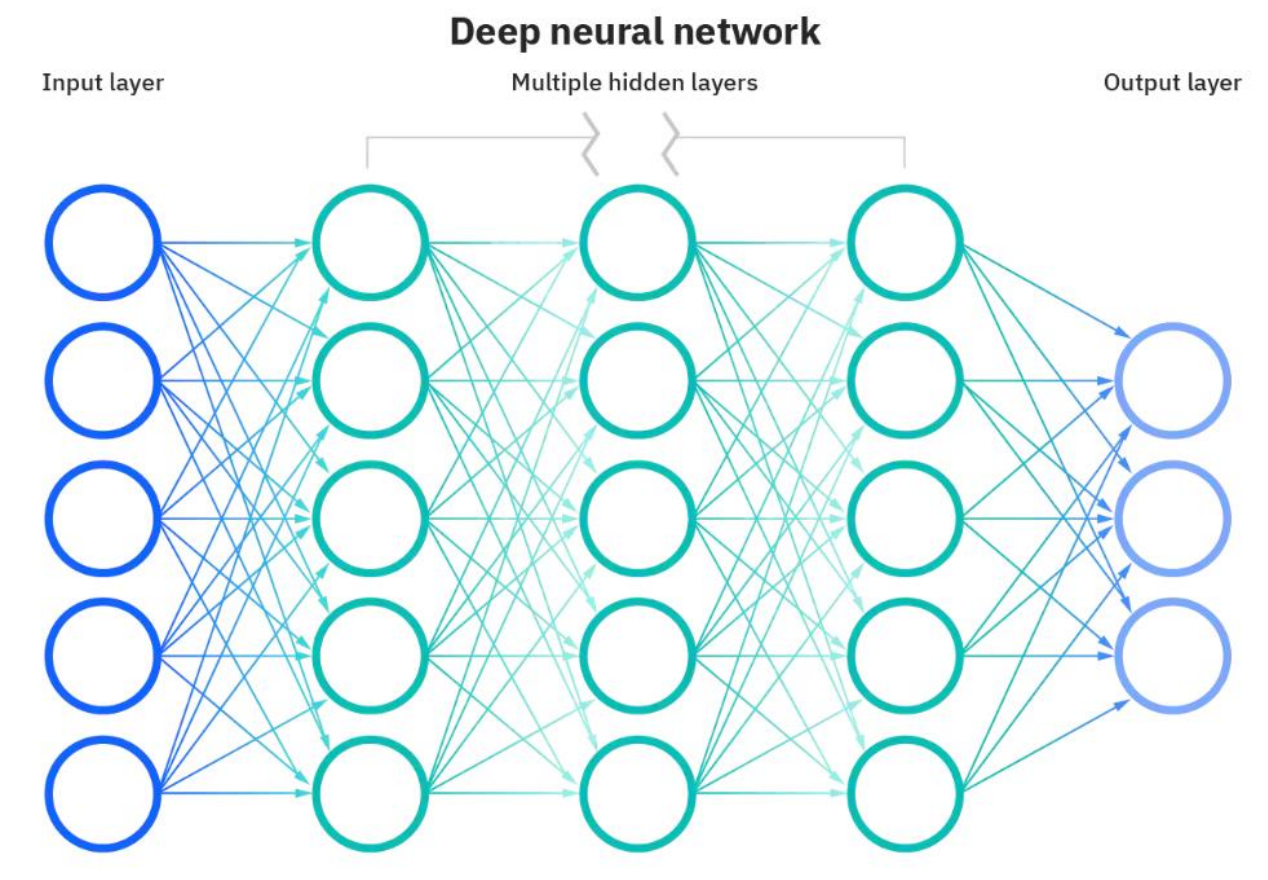
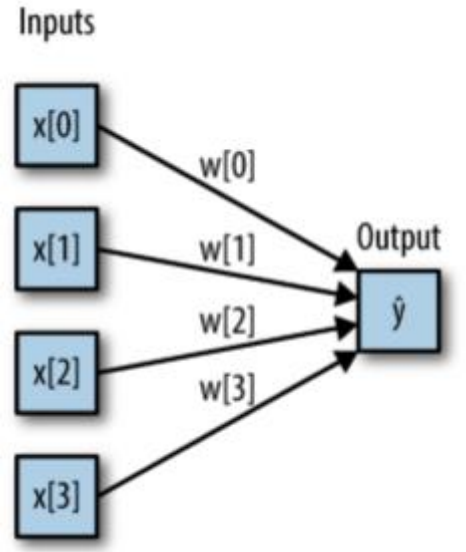
분류 분석 III

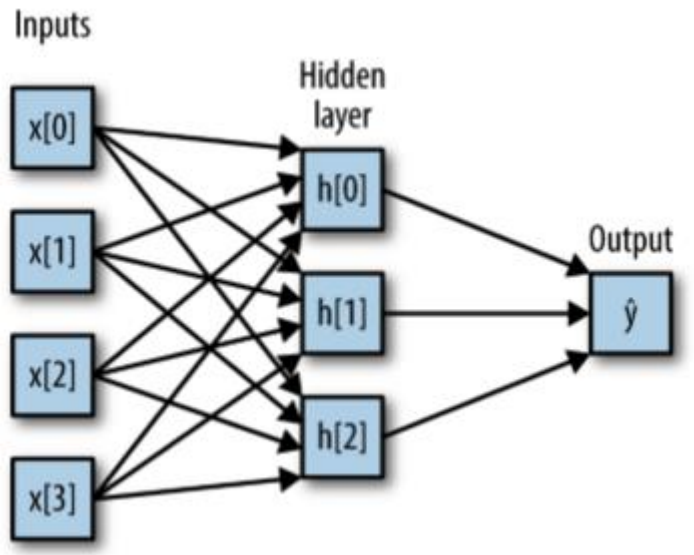
빅데이터 분석

(Artificial) Neural Networks



Linear Regression

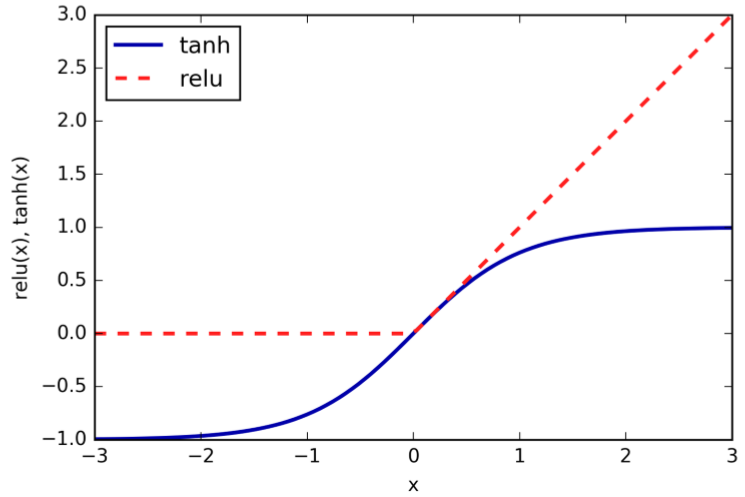


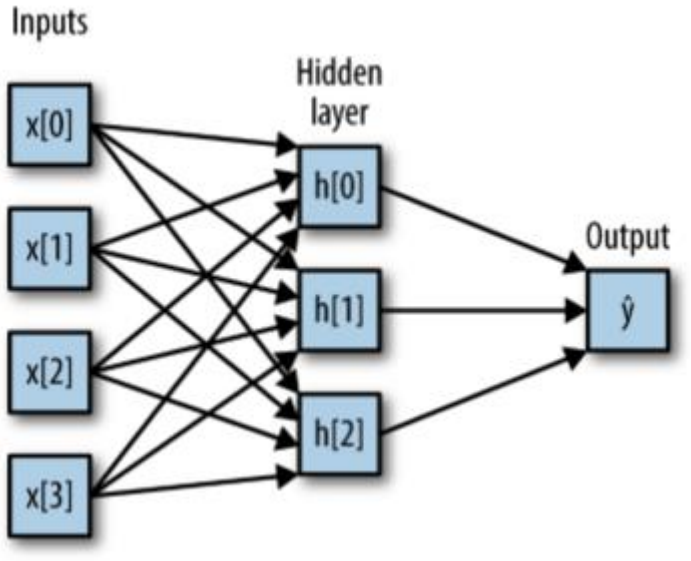


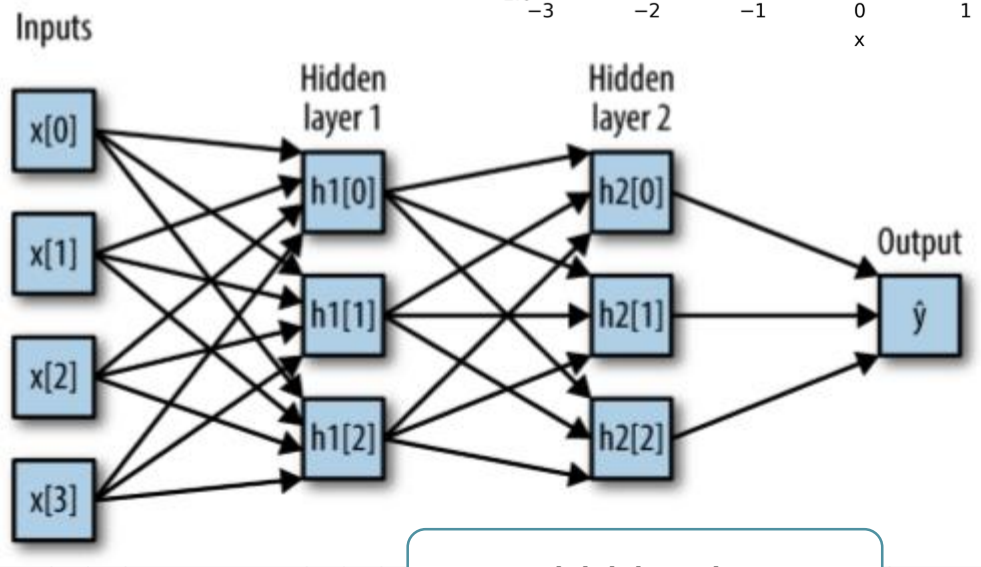
$$\hat{y} = w[0] * x[0] + w[1] * x[1] + \ldots + w[p] * x[p] + b$$

참고문헌 : Introduction to Machine Learning with Python (by Andreas Muller, Sarah Guido)

Multilayer Perceptrons (MLP)







Two hidden layers

$$h[0] = \tanh(w[0,0] * x[0] + w[1,0] * x[1] + w[2,0] * x[2] + w[3,0] * x[3])$$
$$h[1] = \tanh(w[0,0] * x[0] + w[1,0] * x[1] + w[2,0] * x[2] + w[3,0] * x[3])$$
$$h[2] = \tanh(w[0,0] * x[0] + w[1,0] * x[1] + w[2,0] * x[2] + w[3,0] * x[3])$$
$$\hat{y} = v[0] * h[0] + v[1] * h[1] + v[2] * h[2]$$

참고문헌 : Introduction to Machine Learning with Python (by Andreas Muller, Sarah Guido)

붓꽃 데이터

Iris Versicolor

Iris Setosa

Iris Virginica

- **Sepal length** – 꽃받침의 길이 정보
- **Sepal width** – 꽃받침의 너비 정보
- **Petal length** – 꽃잎의 길이 정보
- **Petal width** – 꽃잎의 너비 정보
- Species (**Target**) – 꽃의 종류 정보 (**setosa / versicolor / virginica**)

데이터 로딩 및 데이터 셋 분리

In [1]:

```
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

# 붓꽃 데이터를 로딩
iris_data = load_iris()

# 학습 데이터 셋과 테스트 데이터 셋으로 분리
X_train, X_test, y_train, y_test = train_test_split(iris_data.data, iris_data.target,
test_size=0.3, random_state=11)
```

Neural network 와 Decision tree 정확도 비교

In [2]:

```
# Neural network
from sklearn.neural_network import MLPClassifier

mlp_clf = MLPClassifier(solver='lbfgs', random_state=0, hidden_layer_sizes=[10])
mlp_clf.fit(X_train, y_train)
y_predict = mlp_clf.predict(X_test)

accuracy = accuracy_score(y_test, y_predict)
print('Neural network 예측 정확도: {0:.4f}'.format(accuracy))
```

Neural network

Decision tree

In [3]:

```
# Decision tree
from sklearn.tree import DecisionTreeClassifier

dt_clf = DecisionTreeClassifier(random_state=156)
dt_clf.fit(X_train, y_train)
y_predict = dt_clf.predict(X_test)

accuracy = accuracy_score(y_test, y_predict)
print('Decision Tree 예측 정확도: {0:.4f}'.format(accuracy))
```