IN4320 Machine Learning

Exercise Computational Learning Theory: Boosting

Author:
Milan Niestijl, 4311728

March 12, 2017

a.

Let $x \geq 0$. then $e^{-x} \geq (1-x)$.

Proof:

Note that $e^{-x} \geq 0$ and $e^0 = 1$, so that the result is trivial for $x \geq 1$ and $x = 0$. We assume $x \in (0,1)$. Then we have, using the Taylor series representation of e^{-x}:

$$\begin{aligned} e^{-x} - (1-x) &= \sum_{n=0}^{\infty} \frac{(-x)^n}{n!} - (1-x) \\ &= \sum_{n=2}^{\infty} \frac{(-x)^n}{n!} \\ &= \sum_{n=1}^{\infty} \frac{x^{2n}}{(2n)!} - \frac{x^{(2n+1)}}{(2n+1)!} \\ &= \sum_{n=1}^{\infty} \frac{x^{2n}}{(2n)!} \left(1 - \frac{x}{2n+1}\right) \\ &\geq 0 \end{aligned}$$

Where in the last step it was used that each term of the sum is positive for $x \in (0,1)$.

b.

The implementation of the decision stump is shown below.

```
import numpy as np
import math as m

class DecisionStump():
    def __init__(self, theta=None, feature=None):
        self.theta, self.feature, self.rule, self.accuracy = theta, feature, None, None

    def predict(self, X):
        return np.array([self.rule(x) for x in X[:,self.feature]])

    def score(self,X,y):
        return sum([1 if fxi==y[i] else 0 for i,fxi in enumerate(self.predict(X))])/y.shape[0]

    def fit(self, X, y, w=None):
        m,n = X.shape
            if w is None:
        w = np.ones(m)
        accuracy = lambda x,rule: sum([w[i] if rule(xi)==y[i] else 0 for (i,xi) in enumerate(x)])/sum(w)
        rule1 = lambda theta: lambda xi: 1 if xi<theta else -1 # Can be partially applied
        rule2 = lambda theta: lambda xi: -1 if xi<theta else 1
        rules=[rule1,rule2]
        key = lambda tup:tup[3]
        bestRule = lambda f,rule: sorted([(x,rule(x),f,accuracy(X[:,f],rule(x))) for x in X[:,f]],key=key)[-1]
        bestFeature = lambda f: sorted([bestRule(f,rule) for rule in rules],key=key)[-1]
        self.theta, self.rule, self.feature, self.accuracy = sorted([bestFeature(f) for f in range(0,n)],key=key)[-1]
```

c.

To test the decision stump, data is generated from two normal distributions with the identity as covariance matrix and mean $[0,0]^T, [0,2]^T$ respectively. A scatter plot with the corresponding decision boundary is shown in figure ??.