Exercise Sheet 11

Exercise 1

Implement a `GradBoost` class with a `fit()` and `predict()` method for binary classification problems. The base learner of the `GradBoost` class is sklearn's `DecisionTreeRegressor`. The class should accept the squared and exponential loss as loss function. Implement line search to optimize α_t in each round t.

(a) Test your implementation on a dataset of your choice. Compare your results against sklearn's `GradientBoostingClassifier` with the following parameters:

```
loss='exponential',
criterion='squared_error',
max_depth=1
```

Choose an appropriate learning rate and set the number of estimators as for `GradBoost`.

(b) Plot the dot product $\langle \mathbf{r}_t, \mathbf{h}_t \rangle$, where $\mathbf{r}_t \in \mathbb{R}^n$ is the vector of residuals and $\mathbf{h}_t \in \mathbb{R}^n$ is the vector of predictions of the base learner h_t.

(c) Plot the empirical risk of the base learners h_t as a function of the number t of rounds.

(d) Plot the empirical risk of the ensemble $f^{(t)}$ and the mean $\|\mathbf{r}_t\|_1/n$ of the absolute residuals as a function of the number t of rounds.

Exercise 2

The goal of this exercise is to examine whether Gradient Boosting is prone to overfitting and to investigate the impact of the learning rate.For this, use the digits dataset from sklearn (`sklearn.datasets.load_digits`). Reassign the labels such that digits 0 through 4 are categorized as class -1, and all other digits are categorized as class +1.

(a) Randomly divide the data into a training set (75%) and a test set (25%). Apply sklearn's `GradientBoostingClassifier` using two different base learners:

- Decision Trees with a maximum depth of 1
- Decision Trees with a maximum depth of 3

For each base learner, identify a sufficiently large number of estimators such that Gradient Boosting nearly achieves a perfect training error. Plot both the training and test errors as functions of the number of estimators.

(b) Apply the `GradientBoostingClassifier` with three different base learners using 10-fold cross-validation:

- Decision Trees with a depth of 1
- Decision Trees with a depth of 3
- Decision Trees with a depth of 5

Plot the average test error of each Gradient Boosting variant as a function of the number of estimators.

(c) Apply `GradientBoostingClassifier` with a Decision Tree of depth 5 as the base learner and experiment with different learning rates. Use 10-fold cross-validation for this process. Plot the average test error for each variant of Gradient Boosting as a function of the number of estimators.

Exercise 3

Compare the performance of the following algorithms on the Heart Failure Prediction Dataset (`heart.csv`):

- Logistic Regression
- k-Nearest-Neighbor Classifier
- Decision Tree Classifier
- SVM
- Gradient Boosting
- AdaBoost

Optimize the parameters of each algorithm. Use a holdout-validation as test protocol with split size 0.2 for the test set.

Details on the dataset: [>]