CSC570

Jessica Fong

**Week 7 Breast Cancer (Random Forests) documentation**

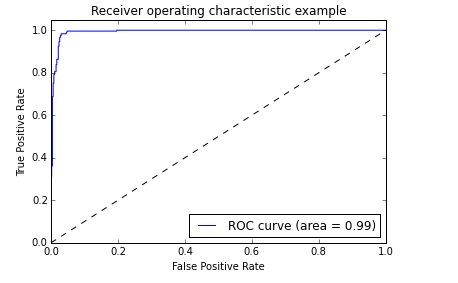
In an attempt to develop a less invasive test for identifying malignant tumors in breast tissue, Chief Oncologist of Seattle Grace Hospital teamed up with a west coast technology company to develop a system that can scan tumors without biopsy. A total number of 699 tumors were examined. The following parameters were measured for each tumor, and their associated malignancy was recorded:

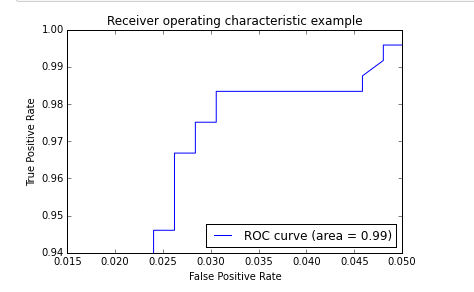
* clump\_thickness
* uniformity\_of\_cell\_size
* uniformity\_of\_cell\_shape
* marginal\_adhesion
* epithelial\_cell\_size
* bare\_nuclei
* bland\_chromatin
* normal\_nucleoli
* mitoses
* malignant

Statistical modeling and analyses have shown that it is possible, based on the given data, to create a machine learning model that can predict if a tumor is malignant or benign using the parameters above.

With an AUC of 0.99232184 (a perfect model has an AUC of 1), the model is likely to:

* predict a true case of cancer (true positive) at a rate of 98.5%
* predict a false case of cancer (false positive) at a rate of 3%
* miss a malignant case at a rate of 1.5%





From the statistical analysis performed, the following plot ranks the parameters from being the most to least important in predicting cancer. The uniformity of cell size is the most important parameter in predicting cancer, along with the uniformity of cell shape and bare nuclei being relatively important.

