Group num : 27
Name/Personal number/Program: ZongHan Hsieh || Nischal Maharjan
Email: zonghan@student.chalmers.se , maharjan@student.chalmers.se
Spent hours : zonghan: 15, Nischal :15
The declaration: We hereby declare that we have both actively participated in solving every exercise. All solutions are entirely our own work, without having taken part of other solutions.

Part1

Summary of The Mythos of Model Interpretability :

This article focuses on supervised learning rather than other ML paradigms such as reinforcement learning and interactive learning. This scope derives from the current primacy of supervised learning in real-world applications and an interest in the common claim that linear models are interpretable while deep neural networks are not. Some papers equate interpretability with understandability or intelligibility, (i.e., you can grasp how the models work). In these papers, understandable models are sometimes called transparent, while incomprehensible models are called black boxes.Like is it converge, unique solution,understand its parameters,complexity.

The demand for interpretability arises when a mismatch occurs between the formal objectives of supervised learning (test set predictive performance) and the real-world costs in a deployment setting. There are some perspectives of models. **Trust**, if the model tends to make mistakes on only those kinds of inputs where humans also make mistakes, and thus is typically accurate whenever humans are accurate. So we need to consider human mistakes. **Causality**, The associations learned by supervised learning algorithms are not guaranteed to reflect causal relationships. There could always be unobserved causes responsible for both associated variables. That by interpreting supervised learning models, you could generate hypotheses that scientists could then test. **Transferability**, transferring learned skills to unfamiliar situations. ML algorithms are already used in these situations, such as when the environment is nonstationary. Models are also deployed in settings where their use might alter the environment, invalidating their future predictions. That means models can be manipulated by giving different induced data. **Informativeness**, while the machine-learning objective might be to reduce error, the real-world purpose is to provide useful information. The most obvious way that a model conveys information is via its outputs. For example, a diagnosis model might provide intuition to a human decision maker by pointing to similar cases in support of a diagnostic decision. **Fair and ethical decision making**, Conventional evaluation metrics such as accuracy or AUC (area under the curve) offer little assurance that ML-based decisions will behave acceptably. Thus, demands for fairness often lead to demands for interpretable models.

There are some points that can be used to observe the interpretability. **Simulatability, Decomposability, Algorithmic transparency, Text explanation, Visualization,Local explanations, Explanation by example.**

Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates

Our takeaways from the papers are listed below

To start off with let’s have a short description on breast cancer. It is a cancer that develops from breast tissue where malignant (cancer) cells form in the tissues of the breast. When a tumor is diagnosed as benign, doctors tend to leave it as it is rather than removing it. As per American Society, when breast cancer is detected early and is in the localized state ,the 5 year relative survival rate is 99% . Problem statement could be predicting a person's tumor (Malignant or Benign) based upon his/her tumor features i.e. its radius, area , smoothness, texture ,perimeter. Machine learning classification techniques are to be used on such types of problem statements.ML takes it as a Probabilistic problem.

Objective here is to use digital image analysis and machine learning to improve breast mass diagnosis based on fine needle aspirates.Digital image analysis coupled with machine learning techniques will improve diagnostic accuracy of breast fine needle aspirates. An interactive computer system evaluates, diagnoses, and determines prognosis based on cytologic features derived from a digital scan of fine-needle aspirate slides. 569 consecutive patients (212 with cancer and 357 with benign masses) provided the data for the diagnostic algorithm. The provided records of patients were histologically confirmed. Data were followed for a year amid any changes in them. Surgical biopsy specimens were taken from all cancers and some benign masses. Patients with the malignant cancer are to receive special treatments. Statistical analysis such as the mean errors were calculated between the actual times of distant disease occurrence and the times predicted using various prognostic features. The predicted diagnostic accuracy was 97% and the actual diagnostic accuracy on 118 new samples was 100%. We can conclude that the Machine learning and computer technology will improve the breast fine needle aspirate accuracy and its prognostic estimation. The subjectivity which is inherent in visual diagnosis can be minimized with computer-based digital image analysis and machine learning techniques . This technology will enhance the usefulness of fine needle aspiration as a diagnostic tool for breast cancer.

Part 2: Implementation

For the first classifier, which is rule based we chose metrics Area, Perimeter and Concavity as a baseline to define whether the patient is benign or not. Initially, the metric chosen was smoothness and texture, which did not produce a differentiating range of value. The values of these metrics for both malignant and benign patients tend to be approx.
So, we chose the other metrics in order to determine the performance of the model classifier. The values are calculated on the provided dataset.
We calculate the mean , minimum and maximum value for these fields and based on the calculated value, we determine how the patient is categorized into the benign and malignant category. Based on the mean value and worst area/concavity, it is calculated. we provide a snippet for the calculations we got

```
Malignant=>
Area:
mean, 978.3764150943397
min, 361.6
max, 2501.0

Perimeter_0:
mean, 115.36537735849062
min, 71.9
max, 188.5

concavity_0:
mean, 0.1607747169811321
min, 0.02398
max, 0.4268
```

```
Benign=>
Area:
mean, 462.79019607843145
min, 143.5
max, 992.1

Perimeter_0:
mean, 78.07540616246497
min, 43.79
max, 114.6

concavity_0:
mean, 0.04605762100840336
min, 0.0
max, 0.4108
```

In the first classifier, the model is built by a given rule and algorithm. We calculate the min,mean,max value to determine the data whether it is abnormal, it’s like an Expert system. It’s perfomance not as good as other ML models, although its conditional value is given by statistical. When we calculate some columns' mean value in the dataset, we found that their value is very similar so we don’t use this column to be our algorithm's conditions. In fact, although their values are very similar, they still have some different things we don’t know. If we consider more conditions it will be more precise. The rule-based model advantage is that it has clear algorithms that let us know how models do, but its disadvantage is that it doesn’t consider all the situations.

In the RF classifier, the model built by a given dataset, finds a good solution to classify whether the data is malignant or not. It’s a common white box solution, which has lots of decision trees in RF. Each tree has a rule that can distinguish the data and vote to the final label. In our experiment, RF got a good value of accuracy, 0.96. RF’s benign recall rate is 0.98, malignant recall rate is 0.94. That means the model miss-find out benign ability is better than malignant.
RF is the decision tree-based random forest classifier. The idea is to reduce variance in the prediction of several noisy decision trees by averaging their results. In terms of interpretability, it is between conventional machine learning models and deep learning. Because of lots of trees, the amount makes it more like black-box. In Ahsan Saeed experiment, they found that at low depth, the trees in a random forest tend to be similar, so we see a high positive correlation. The correlation decreases as the depth increases because the trees start to split on different features. Importantly, we see patterns that

indicate the potential formation of clusters at each level of depth. Also, he proves cluster formation at different levels of depth by using t-SNE. So when we use a RF model, it divides the dataset to numbers of trees, after depth grows the tree will tend to cluster with each other. The key idea is that underlying decision trees in the random forest can exhibit clusters. Once we are confident about the clusters, the last step is to pick a representative unpruned tree from each cluster and present it to the client.

Test data report:

	precision	recall	f1-score	support
0	0.96	0.98	0.97	108
1	0.97	0.94	0.95	63
accuracy			0.96	171
macro avg	0.97	0.96	0.96	171
weighted avg	0.96	0.96	0.96	171

Logistic regression uses all the variables that are fed to it and perform a correlation analysis prior to modeling using a legit model. If correlated variables are not removed, there is a high chance of inflating/ misattributing the impact of a variable on the dependent variable.Logistic regression gives an interpretable model and it runs more quickly with similar accuracy for many problems. Logistic regression has the advantage of being a simple model, easy to train, fast for inference and interpretable and it is a very attractive option for classification in cases that the relation is simple and linear.

	precision	recall	f1-score	support
0	0.96	0.99	0.98	108
1	0.98	0.94	0.96	63
accuracy			0.97	171
macro avg	0.97	0.96	0.97	171
weighted avg	0.97	0.97	0.97	171