Reading and reflection

<u>Machine learning techniques to diagnose breast cancer…</u>

This article was about a constructed classifier that used image processing techniques to assess whether a cancer cell was benign or malignant. They used pictures from fine needle aspirates and used image processing algorithms to gain 30 features for each sample. They used a single-plane classifier to divide data points into benign or malignant and this plane was determined based on mean texture, the worst area, and the worst smoothness. The estimated accuracy of their model was 97% based on tenfold cross-validation. They mentioned in the discussion that this computerised system is intended to address less experienced observers.

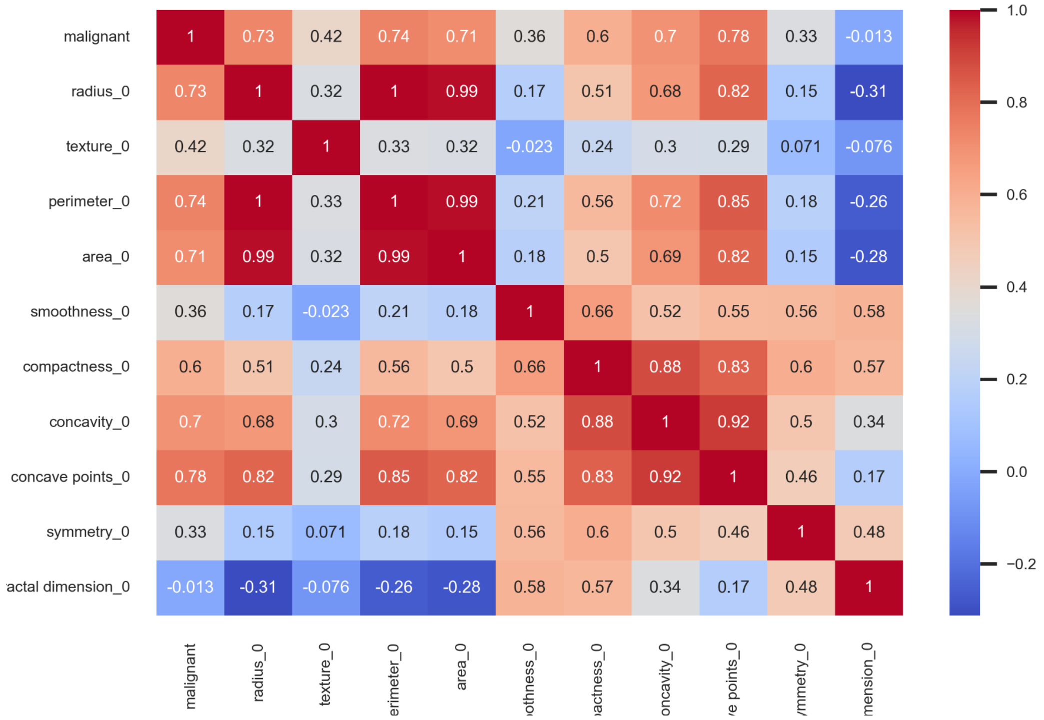
<u>The Mythos of Model Interpretability</u>

This article discusses the importance of interpretability for machine learning systems and the fact that the concept of interpretability is often misunderstood and ill defined within the field. Zachary C. Lipton suggests multiple common ways to model interpretability, such as post-hoc explanations and transparent models while commenting on their flaws. The conclusion is that neither of these approaches provides a perfect solution to the problem of model interpretability. Lipton also discusses the trade off between interpretability and performance since models that are highly accurate often are more complex and thus harder to interpret, while models that are easy to interpret lack in accuracy.

Implementation

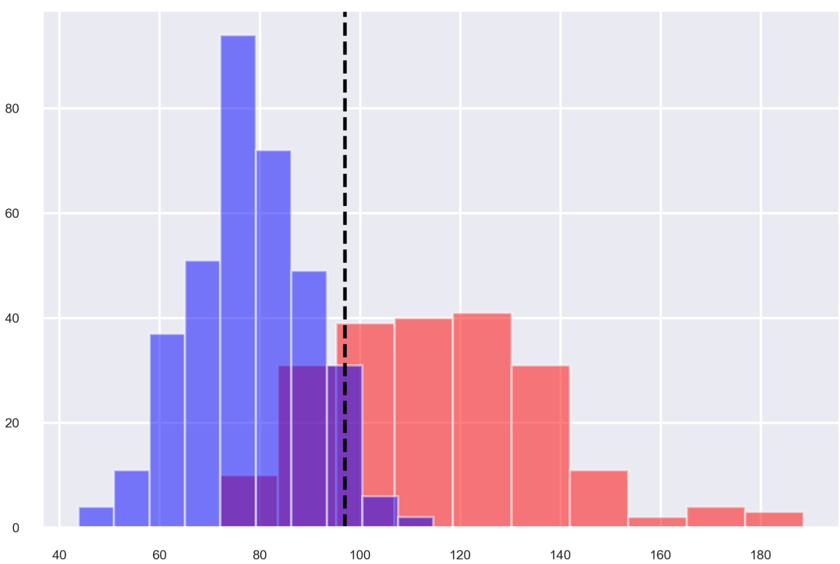
1. Rule-based classifier

We took the approach to find which feature that has something to do with our conditions that has the biggest correlation with malignant.



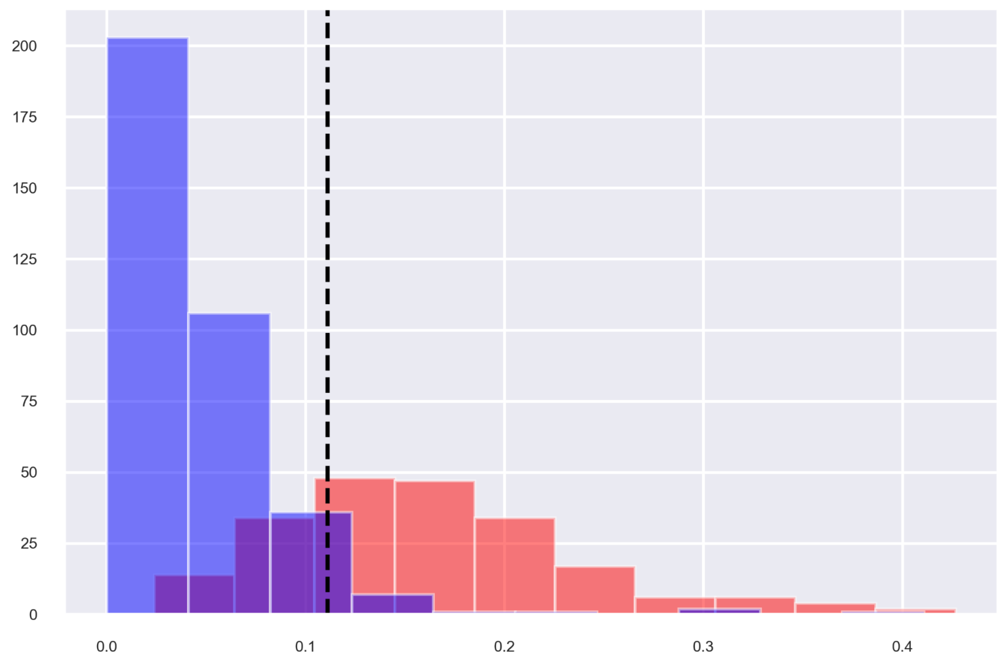
- Cell size

The features which has a connection to size in the dataset are the radius, perimeter and area. The one feature of these three which has highest correlation to malignant was perimeter. The distribution of perimeter for malignant and benign looks like this, red indicating malignant and blue indicating benign:



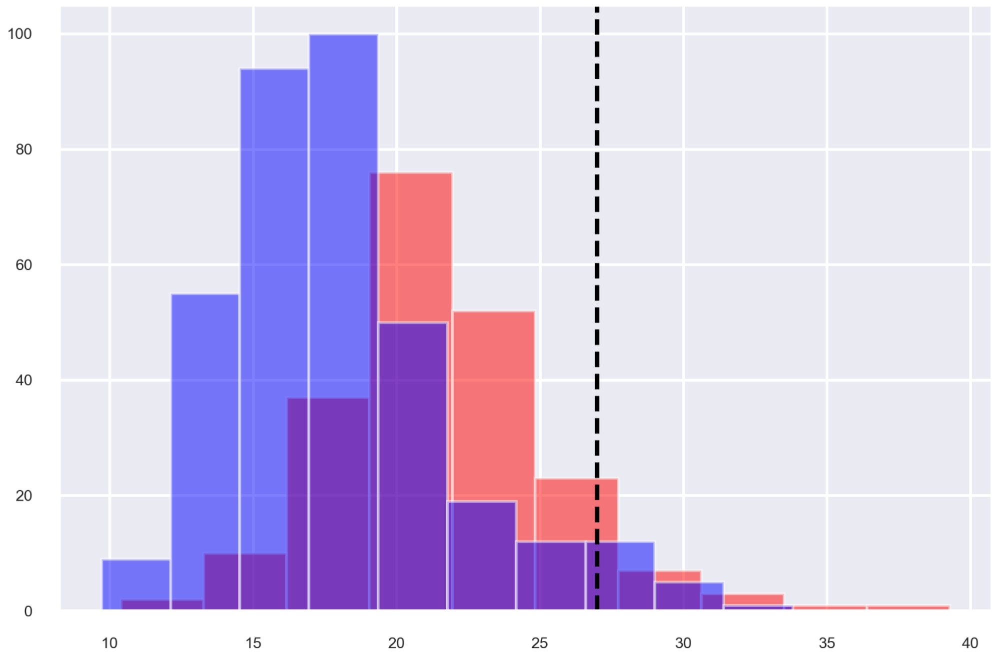
Threshold is at 97.03

- Cell shape



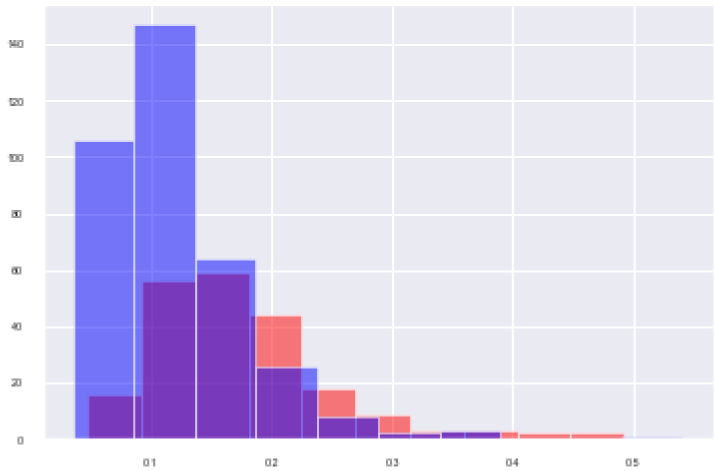
Threshold is at 0.1112

- Cell texture



Threshold is at 26.99

- Cell homogeneity



For the classification using random forest we get the following metrics

Summary - Hanna

<u>Diagnostic systems 14/2 - Hanna</u>

- History of AI in medicine

Before deep learning there were a lot of knowledge based expert systems in AI. Human experts built a large knowledge base and you could do inference based on what was in this knowledge base. An example was the system Mycin which was used to diagnose severe infections so that patients were not treated with antibiotics when it was not needed. INTERNIST was another rule based system for diagnosing over 500 diseases. In the 1990's neural networks started to be used in medicine for diagnosis. They only had a small handful of features. The drawbacks of the early system is that they rely on manually created knowledge bases.

- What is diagnostics

Diagnosis in medicine refers to the task of finding the disease based on some symptoms. However diagnosis goes far beyond medicine as things like debugging and finding out why your car is not working is also considered diagnosis. We have to base our diagnosis on some kind of data. Doctors collect data by asking questions and running tests. You have to go from data to a diagnosis, which can be done (more traditionally) with decision trees. There are different types of diagnosis such as diagnosis made on signs and symptoms, laboratory tests and imaging. A diagnosis test should admit some causes and rule out others. For binary tests there is a tradeoff between sensitivity and specificity when moving the threshold for a positive test result. There are also statistical tests that aim to estimate the probability that the assigned label is the right one. A simple model of the type is Naïve Bays which assumes that all symptoms are independent given the diagnosis.

- Opportunities for machine learning

Almost all data produced in healthcare is digitised which means that we can use the data in machine learning algorithms. There are also large databases that you can licence or use for free such as MIMIC and MarketScan where the latter is bata based on insurance records for 230+ million patients. It has become much easier to work with ML in medicine since the medical records are becoming more standardised. ML in medicine is an important tool to improve consistency in diagnosis and also speed up diagnosis.

- Why is medicine special?

In medicine the stakes are a lot higher than in other areas, making a correct diagnosis can be the difference between life and death. This means that we need robust algorithms, checks and balances to detect errors, fair and accountable algorithms and "interpretable" outputs. Another thing that differentiates the area of medicine is that every action has a cost so we need to build systems that are aware of this.

<u>Follow up 17/2</u>

One thing that rule based translation systems and state of the art neural networks translations have in common is that they both use the concept of interlingua. Sometimes the old school rule based systems can be preferable over the modern. This can, for example, be for niche applications and when not enough data is available. We discussed the implementation of the translation system and a common solution was to translate word for word and then find the best combination of the words using the language model. For Swedish to English this might work pretty well since the order of words are similar in the two languages, but when this is not the case the translations will be faulty.

How would you compute frequencies if you have 10 TB of text? You could take a random sample and compute the frequencies of that insead, but this assumes, of course, that you have a representative sample that is large enough. You could also parallelize the task, but this does not when the steps depend on each other.

We also discussed biases in ML algorithms as the translations of words can reflect stereotypes such as for the case with the translations from Estonian to english. ML algorithms that review resumes can also be biassed since it might learn that some people usually are not hired and then disregard those resumes. For example, there are fewer women working in IT and an algorithm might learn that women are not appropriate for the job.

Summary - Johan

<u>Diagnostic systems 14/2</u>

1. A brief history of AI in medicine

Diagnostic systems started of with a rule-based system. In the 1970's Mycin was an early system developed at Stanford, which performance was estimated to be at par, if not better, than experts. In 1990's neural networks started to be used for diagnostics, but only for a handful or dozens of features. There were some drawbacks of these early methods, such as lack of training data.

2. What is diagnostics

It refers to find the task of finding out the cause of some phenomenon. Diagnostics can be used any many domains, not only healthcare. Such as in software debugging. There are several types of medical diagnosits: clinical, laboratory, radiology. They use different types of data to diagnose. Diagnostic tests should admit some causes and rule out others. There is a problem with diagnostic diseases when it comes to choosing threshold, where data above this threshold gets diagnosed with a certain disease. This is because diseases a normal distrbuted, meaning that at the extremes, some people may not have this disease, even though their data are above this threshold. This means that we have to pick a threshold in regards to sensitivity and specificity tradeoff. Meaning to minimize false positive and false negatives.

Model-based tests could be as simple as assigning labels from a threshold, but there could be a better way to use statistical models and machine learning algorithms to diagnose.

3. Opportunitites for AI

Adoption of electronic health records has increased 9x since 2008. Beyond electronic health records, there exists large datasets. For example, MIMIC and MarketScan. Databases has also becomed more standardized, which makes it easier to extract data for AI. Machine learning / statistics has gotten better algorithms for learning with high-demensional data, there has been huge avvances in image analysis, and better tools (scikit-learn, TensorFlow,...). Big areas of impact of AI in diagnostics is to improve consistency, improve speed, and exploration.

4. Why is medicine special

Compared to other tasks, in medicine, the stakes are higher. This means that we need robust algorithms, checks and balances to detect errors, fair and accountable algorithms, "interpretable" outputs. The systems that we build needs to have rationality in mind, meaning that taking measurement and treating patients has a high cost, so the system needs to give recommendations with these costs in mind. The data in mediceicne is also different. Healthcare data rarely comes with labels, sometimes the number of samples is small (rare diseases), lots of missing data like time intervales for example.

Follow up 17/2

We discussed if there are cases were an old school rule-based system would be preferable than a modern. Richard brought up an example from a clothing store where they wanted more control over the translation of their clothes, because of marketing purposes, and in that case they used rule-based translation models.

Then we moved on to discuss how we approached the decoding task in the assignment. One interesting approach was one group who translated word for word in the sentence, and then swapped the words to see if the translated sentence got higher probability.

We then moved on to discuss how we would compute frequencies if we had to work with 10 TB of text. Someone mentioned to take a random sampe of the data set, so that the sample represents the whole data at large, and using the sample to compute frequencies. Another approach is to use several computers to compute frequencies in pararell.

Next discussing topic was other application areas where we use techniques that we have learned from the previous module. The thought was that we have learned to calculate probabilities of sequences, hence this technique could be used for building a chatbot for example.

We then moved on to discuss questions in the implementation of our assignment.