Portfolio: ML From Scratch

Neo Zhao

CS4375 - Introduction to Machine Learning

- Output of fourRegression.cpp

```
Opening file: titanic_project.csv.
Heading: "","pclass","survived","sex","age"

 ========= Coefficients =========
    Intercept = 0.999875
    "sex" = -2.41086


 ========= Metrics =========
    Accuracy = 0.784553
    Sensitivity = 0.695652
    Specificity = 0.862595


 ========= Run Time =========
    Time: 0.0253963 seconds
```

- Output of fourNaiveBayes.cpp

```
Opening file: titanic_project.csv
Heading: "","pclass","survived","sex","age"

========= Summary =========
    A-priori Probabilities: 0.61 0.39

========= Conditional Probabiliites =========
     Predictor: "pclass"
    0: 0.172131 0.22541 0.602459
    1: 0.416667 0.262821 0.320513

     Predictor: "sex"
    0: 0.159836 0.840164
    1: 0.679487 0.320513

     Predictor: "age"
     Mean & Variance 0: 30.3914 204.73
     Mean & Variance 1: 28.8077 209.316


 ========= Metrics =========
    Accuracy = 0.784553
    Sensitivity = 0.695652
    Specificity = 0.862595


 ========= Run Time =========
    Time: 5.673e-05 seconds
```

- 3(C) - Compare and Contrast Generative Classifiers vs. Discriminative Classifiers
    - Generative Classifiers are in the form of P(Y) or P(X | Y). We can take the estimate parameters straight from these functions and use the Bayes rule to calculate P(Y | X). Naive Bayes and Bayesian Networks are examples of Generative Classifiers.
    - That leads us to Discriminative Classifiers, which take the form of P(Y | X); we may also take the estimated parameters straight from this function. Logistic Regression and Scalar Vector Machines are examples of Discriminative Classifiers.
        - Works Cited

            - Joshi, Prathap Manohar. “Generative VS Discriminative Models - Prathap Manohar Joshi.” *Medium*, 1 Sept. 2018, medium.com/@mlengineer/generative-and-discriminative-models-af5637a66a3.

- 3(D) - Reproducible Research in Machine Learning
    - To define reproducibility, would mean we are able to repeatedly run our algorithm on datasets and obtain the same or similar results of a project. Reproducibility adds importance to each iteration to ensure that we can reduce risks as well as have “deployments for clients become routine.” Furthermore, having iterations that are near perfect creates credibility in the Machine Learning of a project. Replication is key in scientific processes and many researchers have failed to successfully replicate another scientist’s experiments.
    - We can implement successful reproducibility by publishing all data, standardizing experimental protocols, as well as track samples. Transparency to the public allows others, who would like to replicate the project, to know what could go wrong, if it has gone wrong, and how to fix errors. Reproducibility is important. For, if we cannot replicate the experiment, we can’t trust it.
        - Works Cited

            - Ding, Zihao. “5 - Reproducibility.” *Machine Learning Blog | ML@CMU | Carnegie Mellon University*, 24 Aug. 2020, blog.ml.cmu.edu/2020/08/31/5-reproducibility.
            - “The Importance of Reproducibility in Machine Learning Applications.” *DecisivEdge*, 14 Oct. 2020, www.decisivedge.com/blog/the-importance-of-reproducibility-in-machine-learning-applications/#:%7E:text=Reproducibility%20with%20respect%20to%20machine,reporting%2C%20data%20analysis%20and%20interpretation.
            - *Reproducibility: 8 Steps to Make Your Results Reproducible*. blog.genofab.com/increase-reproducibility. Accessed 2 Oct. 2022.