**What’s The Diagnosis?**

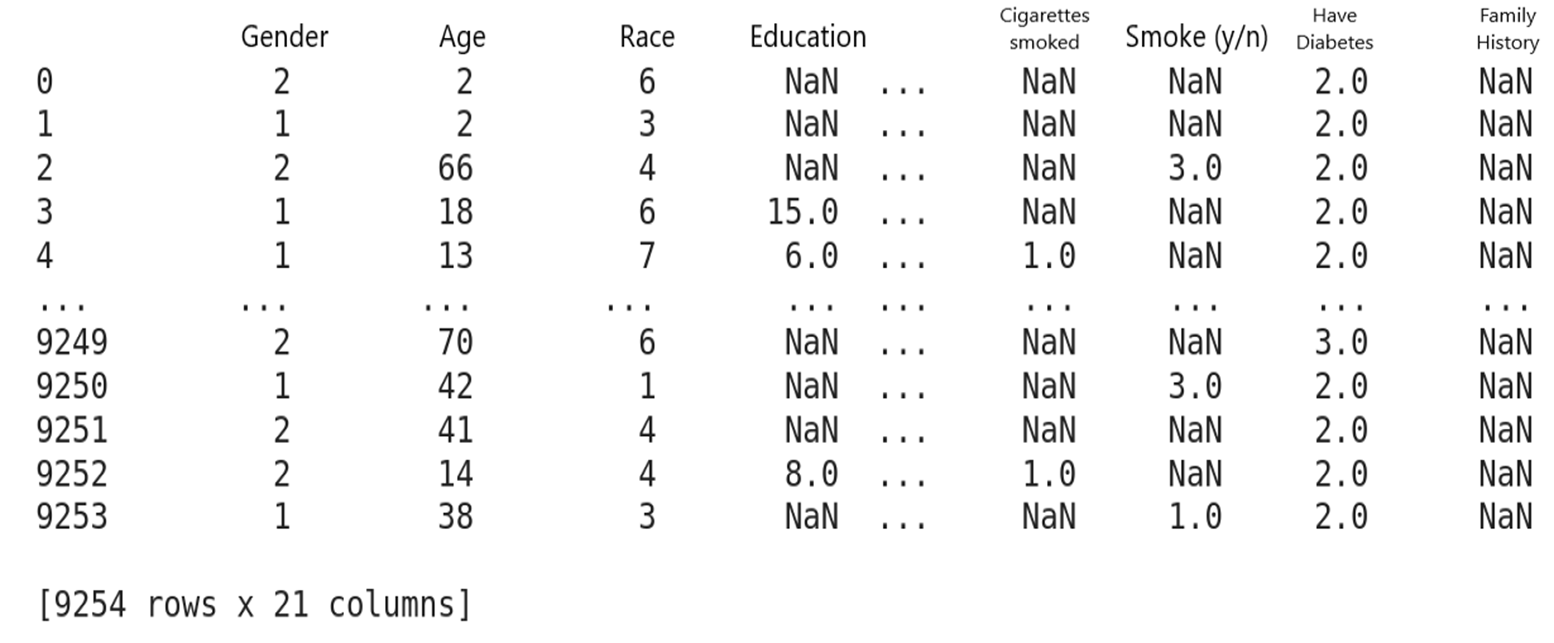
By: Estefania Carrillo, Micheal Myers, Jodh Pannu, Francisco Perales, and Nathan Mauch

**Problem Description**

As the amount of data in the medical world continues to increase and the demand for easy interpretation of such data rises, the need for greater involvement of AI in the medical field to understand and interpret this data becomes increasingly apparent. Healthcare related technologies have increased the availability of medical data, yet the field still largely relies on the discernment of medical professionals to make quick and accurate diagnoses. Meanwhile, the importance of providing timely and efficient patient diagnoses to quality of patient outcomes is highlighted by the ability of early treatment plans to avoid complications associated with certain maladies or even avoid certain diagnoses altogether.

According to the CDC, over 34.2 million Americans currently have diabetes; meanwhile approximately 1 in 3 American adults have been diagnosed with prediabetes. As these numbers continue to rise, the early detection of individuals at risk for diabetes becomes increasingly important. Given the recent advances in AI and the concerning statistics on diabetes in the United States, this project will work automate the assessment of diabetic risk in a patient using an artificial neural network (ANN) to consider various individual health markers and diagnose patients as nondiabetic, diabetic, or borderline. The ANN takes as input relevant patient variables such as race, gender, weight, waist size, BMI, physical activity, etc., before running the inputs through its pattern recognition network to generate a diagnosis of 0, 1, or 2, representative of nondiabetic, diabetic, or borderline diabetic labels. After training, the Shapely Additive Explanation (SHAP) identifies the features most relevant to diagnosis and the generated diagnosis is compared to the known diagnosis to determine accuracy.

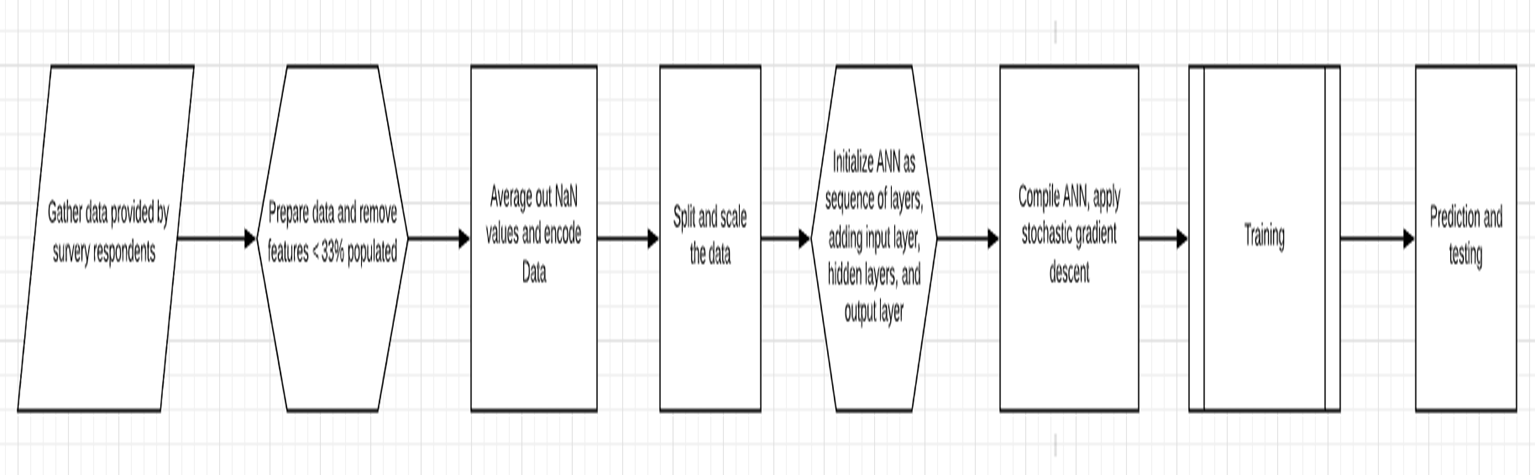
**Framework**

**Data Collection**

The dataset selected for this project includes features from multiple sub-datasets from the CDC’s National Health & Nutrition Examination Survey from 2017-2018. These datasets included questionnaire responses and health status markers from the 9254 individuals surveyed. Across these datasets, our team originally selected 21 different features of interest related to gender, age, race, education, income, height, waist size, BMI, hypertension, physical activity, smoking history, and family history. The datasets were sorted according to participant identification numbers and selected features were combined into one single dataset of interest. Prior to introducing the dataset to the model, the data was then further cleaned as features with a response rate less than 33% were removed and blank or non-existent entries within a sample were filled with the average of the corresponding feature. As such, the final dataset included 9254 samples with 15 different features related to gender, age, race, education, income, height, waist size, BMI, hypertension, and physical activity, that was then divided into testing and training datasets using a 60-40 ratio.

**Approach**

**Overview**

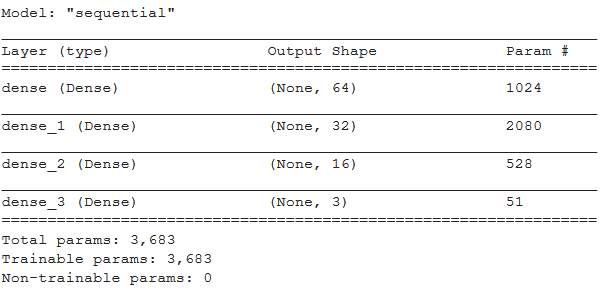
An artificial neural network (ANN) that takes as input relevant patient variables such as race, gender, weight, waist size, BMI, physical activity, etc., will be implemented to have multiple layers connected by a set of adjustable weights that allow signals to travel throughout the network. This network is built of three layers, the input variables given by the selected features of interest, the hidden layer used to detect patterns, and the output diagnosis labeling the samples as 0-nondiabetic, 1-diabetic, or 2- borderline diabetic. Further, a Shapley Additive explanation (SHAP) analysis was then conducted on the trained model to connect optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions. The application of this feature analysis on our model specifically allows us to isolate the features most influential to determining the model’s diagnosis.

**Methods**

In order to build our model, the appropriate framework was first developed through the data collection and cleaning methods described above. This clean dataset was then split and scaled such that the designated training dataset was used as inputs to initialize the ANN. The ANN then applied a stochastic gradient descent and underwent training. After training, the model produced an overall training data accuracy of 90%. This now trained model was then given the remaining testing dataset samples as inputs and used to predict the diagnoses. The diagnosis generated by the model on the testing data were then compared to the samples’ known diagnosis labels to determine accuracy.

**ANN Architecture Used**

We used a sequential dense layer deep neural network architecture. The ANN contained 3 hidden layers, an input layer with 15 inputs and an output layer with 3 outputs for the 3 different classes. The number of trainable parameters was 3.6k. A detailed view of each layer is given in the table below along with the number of parameters per layer.

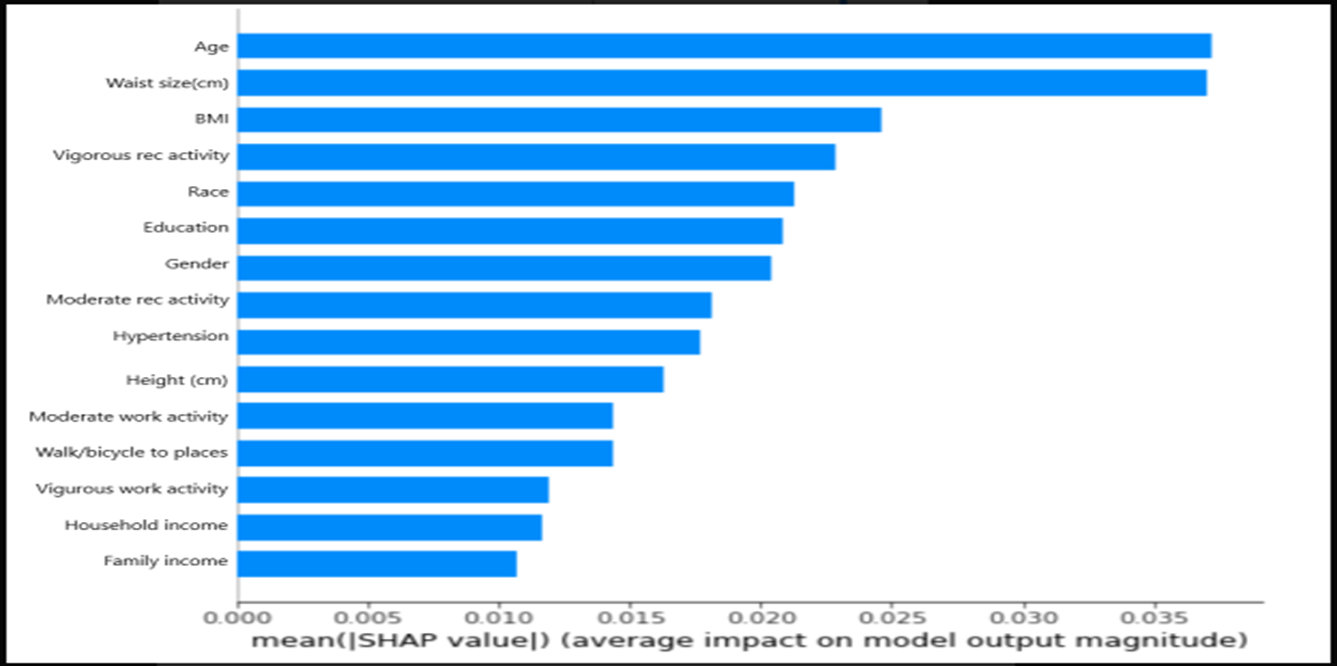
****

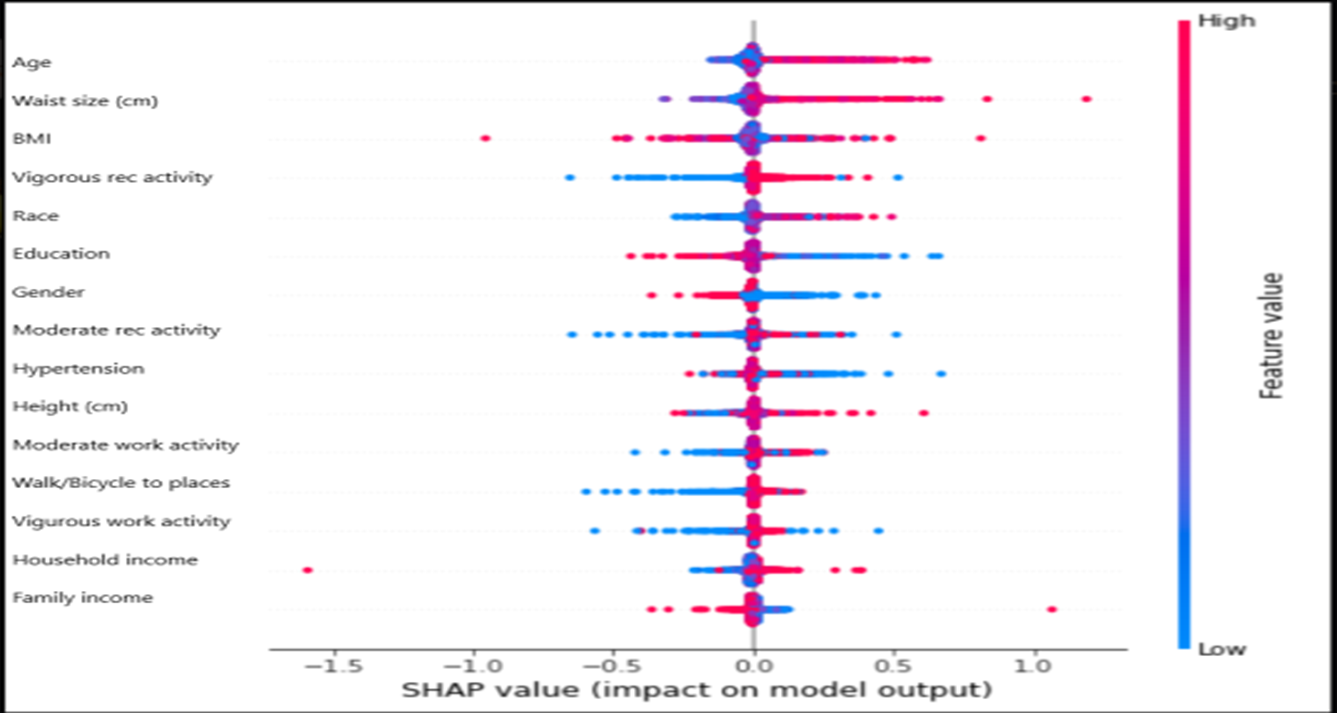
**Attribution Analysis of the Trained ANN**

Attributing the decisions of a neural network to the features of the input has become an important problem. Solving this problem is useful for a variety of reasons including explain-ability to the end-user and increasing network robustness. These important features, also known as attributions [4], can be computed for the entire dataset or individual predictions. Gradients used during the training phase of deep learning have been used to generate saliency maps [5]. Attribution techniques proposed in the literature assign positive and negative importance to each of the input features used by the machine learning model. Attributions can impact both the actual decision of the machine learning model and counterfactuals. Prominent attribution methods calculate the gradient of the prediction function corresponding to the input. Such approaches have been extended using revenue division and cooperative game theory for counterfactual analysis.

We use DeepSHAP [6] to obtain attributions in our project. In DeepSHAP, the attribution values are computed for small components of the neural network, and then combined together recursively. Unlike other methods, DeepSHAP does not heuristically linearize the components. Instead, it uses the attribution values computed for every component element to derive the linearization in an algorithmic manner.

**Results/Error Analysis**

The results of our analysis can be seen in the figures shown below.

In the first figure, the input features of interest are plotted against the respective averaged SHAP values calculated over the entire training dataset. Generally, these values are read such that higher SHAP values correspond to greater importance in the decision making of an ANN.

In the second figure, the red coloring represents a feature’s positive influence on the ANN’s decision, while the blue coloring represents a negative influence. Features with elongated red values are interpreted as applying a positive bias on the category being analysed, in this case diabetic patients; similarly, features with elongated blue values are understood to apply a negative bias in the same category. As such, this selective coloring serves to explain the decisions made by the ANN. In the case of our model, the most significant positive biases include age, waist size, and BMI, while features such as income and work activity were among the least relevant.

**Error Analysis**

Error in an artificial neural network can come from several sources that may be either human error or systematic. Was information mislabeled, was information missing, or is it that more layers should have been added or changed some of my parameters. With a study that involves such a complex diagnosis as diabetes that does still have no clear defined cause can be exceptionally difficult to predict and have accurate results. The biggest issue that our study ran into was dealing with missing data, even with the extensive data set and all of the features we had available to us. We did not have every feature for every patient surveyed which could have been a crucial component in making an accurate prediction. The National Health and Nutrition Examination Survey is unique as the data is a combination of interviews and physical examinations. Because of this method of collecting data, you will have different physical examinations take in different sets of data so you could have key fields such as waist size or recreational activity not be present. Another shortcoming of the data set used for our study is hoping that the patients answered their questions fully and honestly. Patients may be embarrassed to admit their physical activity levels or they may not know their family history. These features missing or possibly not being correct may have caused our method to weigh certain features higher or lower.

To improve our model and our method, we would want to learn more on how to better clean the data and reduce features that have very little impact. Also ensure that all of the key features are collected properly and with some form of standardization. The standardization can come from the questionnaire removing ambiguity and questions up for personal interpretation, as well during physical examinations that the medical professionals interpret answers in a more similar manner. The model used was able to do a good job at taking in a large number of features and making good predictions. Until the data could be cleaned better and key features are gathered for each patient, the model should not be changed.

**Review of Literature**

**Support Vector Machine in Diagnosis**

There are several different approaches to using Artificial Intelligence in helping predict diagnosis of diseases, specifically diabetes. The support vector machine learning method has already proven itself to have a high rate of success in the biomedical fields, especially in bioinformatics. This model works by being very data driven and being model-free, which provides a key element of discrimination in classification of the diagnosis, especially in cases where the amount of samples are limited but there are a large number of variables involved. Which in a world such as health care where there is an overabundance of data, especially data that may not prove to be crucial to determining a result such as is a person diabetic, pre-diabetic, or not diabetic. An algorithm that is able to not have to be model driven, such as logistic regression, can be proven to be very helpful in detecting diseases.

In the United States diabetes affects roughly 23.6 million people, which roughly one third of those people are completely unaware they are diabetic. Roughly another 57 million Americans are considered to be pre-diabetic and are at a high risk of developing diabetes, heart disease, and stroke. The results of the supervised machine learning method show that based on using the variables family history, age, race and ethnicity, weight, height, waist size, body mass index, hypertension, sex, and physical activity had either an 83.5% and 73.2% are under the receiver operating characteristic curve. The two characteris schemes were diagnosed or undiagnosed diabetes vs. pre-diabetes or no diabetes and undiagnosed diabetes or pre-diabetes vs. no diabetes. With these kinds of high success results using only characteristics that are gathered in a routine clinical visit and no lab results needed, health care providers would be able to treat their patients a lot sooner and suggest lifestyle changes to prevent that person becoming diabetic where early screening and diagnosis are crucial with this disease.

The SVM model proved to be a very effective technique to be able to detect cases of diabetes and pre-diabetes in the US. It demonstrated that the discriminative performance was equivalently as effective as the already commonly used epidemiological method of multivariate logistic regression. It also showed to be just as effective in detecting a diagnosis without any lab results. This model and test went a long way to prove that complex diseases could effectively be diagnosed with simple variables and could easily be extended to different populations outside of the US or for detecting different diseases.

**Artificial Neural Network for Diagnosing Diabetes:**

An artificial neural network is a way of processing data that can be considered similar to a human brain. The processing of data is done by the interconnected layers that work in tandem to solve a complex problem such as diagnosing a disease. This is the neuronal structure. After that is done the layers learn by creating a network between these neurons and use a learning algorithm. Typically there are three layers in the neural network, the input layer where the raw clinical data enters the network, the hidden layers whose performance is determined by the inputs and the weight in the relation between the inputs, then finally the output layer which is hidden depending on the activity of the unit and the weight of the connection between the hidden and output unit.

Along with the SVM model, an ANN proved to be a very effective way of being able to diagnose diabetes in patients with diabetes with simple variables. Even without knowing the precise cause of type I diabetes, issues that increase the risk of developing the disease range from family history, environmental factors, harmful immune system cells, geography, race, age, and blood pressure. A lot of this information is collected in everyday clinical situations and can be input into a neural network to let a health care professional know that a patient may be at risk of diabetes. This was even the system we used in our experiment which boasted an 83% accuracy rating and a similar study done using ANN boasted an even better 87.3% prediction accuracy.

A huge advantage that an ANN would have over other practical approaches such as regression analysis is not having to fit the data into an appropriate function to have all the collected data and adjust the output automatically when more information is attained. The artificial neural network, which strives to emulate human thinking is an adaptive system that is able to handle new inputs and variables being inserted into the network and making predictions based on the results of those inputs.

**Conclusion**

Ultimately, the ANN created was able to correctly assess 83% of test samples for diabetes; meanwhile, feature analysis distinguished attributes such as age, waist size, and BMI as more influential to generating a diagnosis when compared to less-relevant features such as work activity or income. While the testing dataset’s accuracy rate was ultimately lower than the training dataset’s 90% accuracy, the automated diagnosis still proved to be sufficiently accurate for the majority of samples, suggesting the possibility of its medically-supervised integration to existing data management systems for the auto-assessment of patients’ diabetic risk. Alternatively, the results pose potential for the model’s extrapolation to the development of similar tools for the assisted diagnosis of alternate conditions. Lastly, within the realm of the existing diabetic risk model, potential for future work includes the variation of network samples or features to assess for the diabetic risk of population subset or the relevance of additional/ alternate health parameters to a diabetic diagnosis.

As the amount of data in healthcare continues to increase and ongoing developments in AI produce agents increasingly efficient in data management and the execution of higher-level reasoning, the integration of AI to healthcare could soon prove to become an extremely useful tool. While the realm of AI is not expected to reach the levels of independent competency necessary for replacing health professionals altogether, the streamlined patient health screening possesses the capability to allow our healthcare system to function at higher levels of efficiency. Moreover, this efficiency presents not only the economic benefits, but offers multi-leveled social benefits associated with the early detection and/or early treatment of potentially life threatening diseases and the maintenance of a healthier population as a whole.

**References**

CDC, “National Diabetes Statistics Report, 2020,” Centers for Disease Control and Prevention, Feb. 11, 2020. https://www.cdc.gov/diabetes/library/features/diabetes-stat-report.html (accessed Nov. 24, 2020).

El\_Jerjawi, N. S., &amp; Abu-Naser, S. S. (2018). Diabetes Prediction Using Artificial Neural

Network. International Journal of Advanced Science and Technology, 121, 55-64. doi:10.14257/ijast.2018.121.05

Yu, W., Liu, T., Valdez, R., Gwinn, M., &amp; Khoury, M. J. (2010). Application of support vector

machine modeling for prediction of common diseases: The case of diabetes and

pre-diabetes. BMC Medical Informatics and Decision Making, 10(1).

doi:10.1186/1472-6947-10-16

M. Sundararajan, A. Taly, and Q. Yan, “Axiomatic attribution for deep net-works,” inProceedings of the 34th International Conference on MachineLearning - Volume 70, ser. ICML’17. JMLR.org, 2017, pp. 3319–3328.

K. Simonyan, A. Vedaldi, and A. Zisserman, “Deep inside convolutional networks: Visualising image classification models and saliency maps,”arXiv preprint arXiv:1312.6034, 2013.

S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” inAdvances in Neural Information Processing Systems30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus,S. Vishwanathan, and R. Garnett, Eds.Curran Associates, Inc., 2017,pp. 4765–4774.