CS230

Predicting Small Molecule Binding Affinity to Serotonin Receptors

Annie Marsden*
Department of Computer Science
Stanford University
marsden@stanford.edu

Abstract

We construct a deep neural network to predict the binding affinity to serotonin receptors.Understanding the binding behavior of these receptors is important for a diverse range of future mental health drug research. We show representing small molecules as vectors with binding constants for other targets can provide predictive information. Our neural network approach outperforms a simple linear regression model as well as a more complex matrix completion method. The metric used is the F1 score. The code to reproduce this work is available on github [1].

1 Introduction

Pharmaceutical drugs often work by inhibiting the activity of an undesirable enzyme. Thus the problem of developing a new drug generally involves discovering a small molecule that targets this enzyme and inhibits its activity. The process to do so can be very time-consuming and expensive as it involves empirically testing a large number of small molecules' binding affinities to the target enzyme. A successful application of deep learning methodology to the problem of predicting a small molecule's binding affinity to a given target enzyme would reduce the cost and time required for drug discovery. [2] In this work we construct a deep neural network to predict the binding affinity to serotonin receptors. Serotonin is one of the oldest molecules known to signal behavioral changes across a broad range of organisms– from flatworms to humans [3]. Serotonin is important in regulating our normal physiological functioning as well as our behavior, mood, and cognition. The role of serotonin in the function of the brain was first discovered in the 1940s when lysergic acid diethylamide (LSD) was found to have an enormous impact on human behavior and the structure of LSD was found to be the scaffold for serotonin [3]. Serotonin receptors describe a class of proteins that sit on the membrane of the postsynaptic neuron and, once bound to serotonin, transmit the neural signals responsible for the effects of serotonin. Understanding the binding behavior of these receptors is important for a diverse range of future mental health drug research.

2 Related work

There is not necessarily a large body of work focusing specifically on the prediction of binding to serotonin receptors, however there is a wealth of research done on deep learning in drug discovery methods. One of the main problems in this work is the class imbalance– there are very few positive examples. In [2] the authors demonstrate how one-shot learning can be used to lower the amount

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

of data required to make meaningful predictions in drug discovery when there are few positive examples. They introduce a new architecture based on the well-known long short-term memory architecture and show significant improvement in learning. Another question for this area of research is how to use promising architectures like convolutional neural networks efficiently in the drug-discovery regime. In [4], the authors demonstrate how to apply the convolutional concepts of feature locality and hierarchical composition to model bioactiviy and chemical interactions. Extending beyond convolutional neural networks, in [5] the authors present the use of recursive neural netowrk approaches in the problem of predicting solubility of drug-like molecules. They show competitive performance with other state-of-the-art methods.

3 Dataset and Features

The input of this neural network is a vector representation of a small molecule and the output gives the probability that this small molecule will bind to a serotonin receptor. The vector representation of the small molecule is one of the most important aspects of this work. To represent the small molecule we construct a vector in which the i^{th} entry gives the binding constant, K_B of the small molecule to the i^{th} chosen target. The K_B value of a small molecule/target pair represents the inverse concentration of the small molecule needed to bind to the target. Thus a large K_B value indicates that the small molecule binds well to the target. Here, 939 targets aside from serotonin receptors are chosen to construct these vector representations and $7,139$ small molecules are considered. Not every target/molecule pair has an empirical measurement– for the missing data we assume a K_B value of 0 with the justification that it is very unlikely for a random small molecule/target pair to have any binding affinity. All the data comes from [6].

4 Methods

We use a k-layer neural network with ReLU layers. For the j^{th} layer of the network we have activation:

$$A_j = \max(0, W_j A_{j-1} + b_j).$$

For the first layer we have $A_0 = X$ and for the output of the last layer we have $\hat{y} \in \mathbb{R}^m$ where m is the number of training examples and

$$\hat{y} = A_k = \sigma(W_k A_{k-1} + b_k)$$

where σ represents the sigmoid function.
The loss function we use is

$$L(p) = -\frac{1}{N}\sum_{i=1}^{N} s y_i \log(p_i) + (1 - y_i)\log(1 - p_i).$$

Note we include "scaling factor" s to deal with the fact that there are very few positive examples, thus for larger s we upweight the importance of correctly catching positive examples.
Finally, prediction is done by: $\hat{y}_i = p_i > 0.5$.

5 Experiments/Results/Discussion

The hyperparameters in this model were: $1)$ the number of layers in the neural network, $2)$ the dimension of the hidden layers, $3)$ the learning rate/method, and $4)$ the scaling factor s value. For $3)$ we simply use Adam and set the learning rate to be small enough for smooth behavior. For the rest we conduct a hyper parameter grid search (rather than grid search, as is suggested in class we conduct a randomized search). Our metric of evaluation is the F_1 score. This is defined as,

$$F_1 = \frac{\text{2\# true positive}}{\text{2\# true positive } + \text{\# false negative } + \text{\# false positive}}.$$

For the scaling factor we compare the performance with linear regression, a 1-hidden layer model, and a 2-hidden layer model. In Figure 5 we see that the 1-hidden layer model performs the best and that the optimal scaling factor is in the range $10 - 100$. In the 1-hidden layer model we explore the

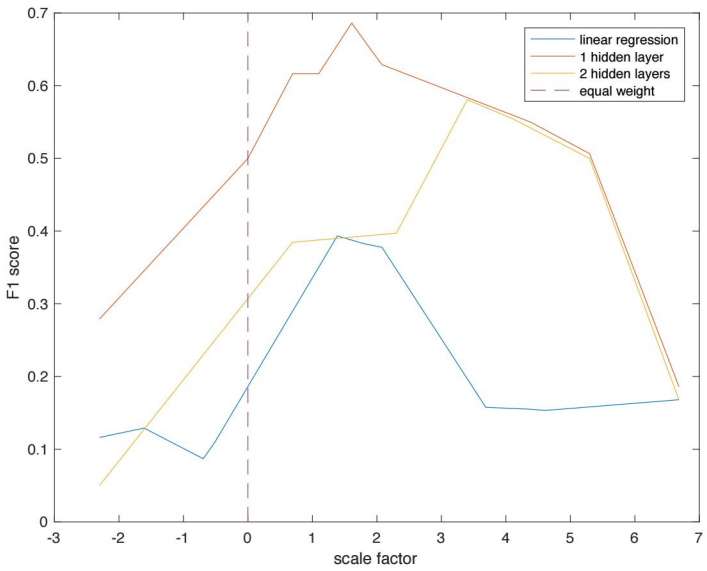

Figure 1: On the horizontal axis is the log of the scaling factor. We see that increasing scaling improves performance for all models.

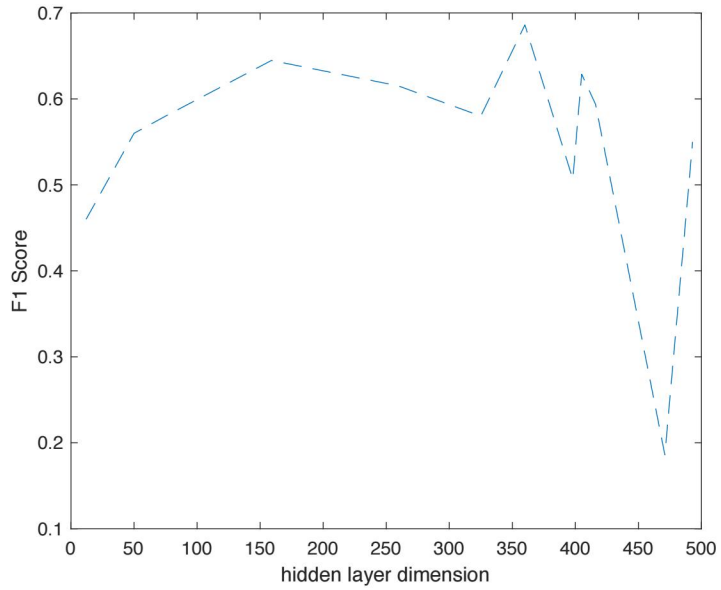

Figure 2: For a 1-hidden layer model we find that the best dimension is roughly 360.

effect of the dimension of the hidden layer and find that the best option is a dimension around 360 so that the dimension of the layer progression is $939 \to 360 \to 1$.

Finally, we compare our neural network performance to a simple linear regression model and a matrix completion model. In table 5 we see that our neural network achieves the best F_1 score and good accuracy on both the positive and negative examples.

Method	F_1 **Score**	**Corresponding P/N Accuracy**	**Increased** s **P/N accuracy**
Linear Regression	0.423	0.46/0.98	0.84/0.38
Matrix Completion	0.571	0.8/0.88	0.8/0.88
Neural Network	0.686	0.66/0.98	0.77/0.89

Table 1: Comparison of 1-hidden layer neural network performance to Linear Regression and Matrix Completion. "Corresponding P/N Accuracy" gives a pair x/y where x is the accuracy on the positive examples and y is the accuracy of the negative examples using the value of s that achieves optimal F_1 score. Note that this does not apply to matrix completion since s is not a hyperparameter here.

6 Conclusion/Future Work

In this work we find that a shallow neural network architecture shows promising performance in predicting the binding affinity to serotonin receptors. Unfortunately the size of the training data is small enough that even a 2-layer neural network seems to be able to overfit. In future work if a larger dataset is attainable it would be interesting to see if better performance can be achieved with a deeper net.

Futhermore, in the literature there are very few baselines for comparison. Future research must implement more comparison methods to get a better sense of the kind of F1 score that would be helpful. It may be possible that research done by pharmaceutical companies can achieve something much much better, however unfortunately this research is not generally publicly available.

We find that, unlike matrix completion methods, constructing a loss function with scaling factor s allows freedom in tuning the specificity/sensitivity ratio. To increase speed of training, dimensionality reduction methods should be explored in the vector representation of the small molecule. Finally, the input data representation here could be adjusted. For instance one could input an image of the chemical structure appended to the input used here. This may provide much more generalizable information.

References

[1] https://github.com/acmarsden/cs230project.

[2] Han Altae-Tran, Bharath Ramsundar, Aneesh S Pappu, and Vijay Pande. Low data drug discovery with one-shot learning. *ACS central science*, 3(4):283–293, 2017.

[3] David E Nichols and Charles D Nichols. Serotonin receptors. *Chemical reviews*, 108(5):1614–1641, 2008.

[4] Izhar Wallach, Michael Dzamba, and Abraham Heifets. Atomnet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv preprint arXiv:1510.02855*, 2015.

[5] Alessandro Lusci, Gianluca Pollastri, and Pierre Baldi. Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules. *Journal of chemical information and modeling*, 53(7):1563–1575, 2013.

[6] The binding database.