Classifiers versus discriminative classifiers.

Generative classifiers and discriminative classifiers are two types of machine learning algorithms used for classification tasks. The main difference between them lies in their approach to modeling the underlying probability distribution of the data. Generative classifiers aim to model the joint probability distribution of the data and the class labels, while discriminative classifiers aim to model the conditional probability distribution of the class labels given the data.
Generative classifiers, such as Naive Bayes, build a model of the probability distribution of the input features and the classes, and then use Bayes' theorem to predict the probability of each class given the input features. Discriminative classifiers, such as logistic regression or support vector machines, do not model the probability distribution of the input features and the classes, but instead learn a decision boundary between the different classes directly from the data.
The choice between generative and discriminative classifiers depends on the specific task and the available data. In general, discriminative classifiers tend to perform better when the number of features is large and the data is well-separated, while generative classifiers tend to perform better when the number of training examples is small and the distribution of the data is complex.

Reproducible research in machine learning

Reproducible research in machine learning is the practice of making research results and experiments transparent, verifiable, and reusable by other researchers. Reproducibility is important for ensuring the validity and reliability of scientific findings, and for enabling the scientific community to build on and extend existing work.
There are several ways to achieve reproducibility in machine learning research. One way is to use open-source software and publicly available datasets, which allows other researchers to replicate experiments and verify results. Another way is to use tools such as Docker or virtual environments to ensure that the same software environment is used across different experiments. Finally, documenting experiments thoroughly, including details about the data, preprocessing steps, hyperparameters, and evaluation metrics, can help other researchers understand and reproduce the results.

Sources:

- Ng, A. (2002). Discriminative vs. generative classifiers: A comparison. In Proceedings of the 2002 Conference on Advances in Neural Information Processing Systems (pp. 841-848).
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
- Stodden, V., LeVeque, R. J., & Mitchell, I. M. (2018). Reproducibility in computational science. Science, 361(6405), 1390-1391.
- Peng, R. D. (2011). Reproducible research in computational science. Science, 334(6060), 1226-1227.