CSC311 Summaries

Wilson Hsu

April 2020

1 Regression Shrinkage and Selection Via Lasso

In "Regression Shrinkage and Selection Via Lasso" by Robert Tibshirani, he proposed a new method called "lasso" that can be used for shrinkage and selection for regression and generalized regression problems. It takes the good features from both subset selection and ridge regression, where it sets some coefficients to 0 to have an interpretable model and shrinks some coefficients to improve stability. Lasso shrinks the coefficients and set some to 0, as it translates the coefficients by a constant factor, and truncating at 0, instead of scaling them by a constant factor like ridge regression. Standard error for Lasso can be obtained via the bootstrap method. Many examples given in the paper shows that Lasso is a good method to adapt compare to subset selection and ridge regression, where it performs considerably better.

Suggestion Usage:
1) Lasso can be used to find features that are important, with its more interpretable characteristic.
2) Can use Lasso to estimate a financial market.

2 The Tradeoffs of Large Scale Learning

In "The Trade-offs of Large Scale Learning" by Leon Bottou and Olivier Bousquet, the trade-offs between approximation, estimation, and optimization errors in large scale learning problems are studied. Large scale learning problems are problem that are constraint by computation time and nothing else. As typically in a learning problem when number of objective functions to choose from increase, we would find error of approximation to decrease, error of estimation to increase, computation time to increase. When number of samples increase, we would find error of estimation to decrease and computation time to increase. When optimization tolerance increase, we would find error of optimization to increase. In this paper, the authors show the fact that generalization properties of large-scale learning systems depend on both the statistical properties of the objective function and the computational properties of the optimization algorithm with four difference gradient descent function.

Suggestion Usage:
1) We can use approximate optimization on tasks that take too long to train, this way we can process more training examples and generalize better.
2) This paper also suggest that we choose a family of objective functions with better statistical properties, so we can generalize better with our large-scale learning system.

3 Imagenet classification with deep convolutional neural networks

In "Imagenet classification with deep convolutional neural networks" by Alex Krizhevsky, Illya Sutskever, and Geoffrey E. Hinton, they discussed how they trained a large, deep convolutional neural network to classify 1.2 million high-resolution images into thousand different classes. The neural network consists eight learned layers, of which five are convolutional layers and three are fully connected layers. On a dataset this big, it is important to speed up the training time, therefore they applied ReLU non-linearity to every learned layer instead of the traditional saturating neuron models. Another important point for a neural network is to prevent overfitting, they combat this problem with the use of a regularization method called "dropout" and data augmentation. This neural network achieved top-1 and top-5 error rates of 37.5% and 17.0% in ImageNet LSVRC-2010 contest and a winning top-5 error rate of 15.3% in ILSVRC-2012 competition. They mention that the depth of this convolutional neural network

is also very important, as removing any middle layers would result in loss for the top-1 performance.

Suggestion Usage:
1) Automatic face detection.
2) Photo search engine (like google image search).

4 Equality of opportunity in supervised learning

In "Equality of opportunity in supervised learning" by Moritz Hardt, Eric Price, and Nathan Srebro, they propose a new way to exclude discrimination from the perspective of supervised learning. The ideology is called "Equal opportunity" where the predictor predicting true is conditionally independent with the protected attributes on true target. This idea unlike demographic parity, allows the predictor to depend on protected attributes but only through the true target variable. This encourages the use of features that relate to true target directly and not through the protected attributes. Equal opportunity is a weaker notion of non-discrimination, but that is why it can allow for better utility. In the paper, a Bayes optimal predictor is derived and a case study on FICO scores is conducted. An important feature of this proposal is that it can be achieved via a simple and efficient post-processing step, which requires only aggregating information about the data and therefore could be carried out in a privacy-preserving manner.

Suggestion Usage:
1) Automatic immigrant application screening, as these should definitely be evaluated with no discrimination.
2) Insurance company or financial institution's credit score calculation.

5 Human-level control through deep reinforcement learning

In "Human-level control through deep reinforcement learning" by Volodymyr Mnih, Koray Kavukcuoglu, David Silver, et al, they created a deep Q-network, which combines reinforcement learning with deep neural network that would be able to solve a wide range of challenging tasks. Reinforcement learning is known to be unstable when a nonlinear function approximator is used to represent action value function such as neural network. As a solution, authors address these instabilities with an efficient variant of Q-learning that removes correlations in the observation sequence, smoothens over changes in the data distribution, and reduces correlations with the target. The algorithm is tested on 49 Atari games, where it outperforms the best reinforcement learning methods at the time on 43 of the games, and achieves higher score than 75% of the human player on 29 games. It is shown that this single deep Q-network is able to learn a variety of control policies in many different environment with the same set ups.

Suggestion Usage:
1) Can be applied AI player in board games like Go, chess.
2) Stock market predictions.