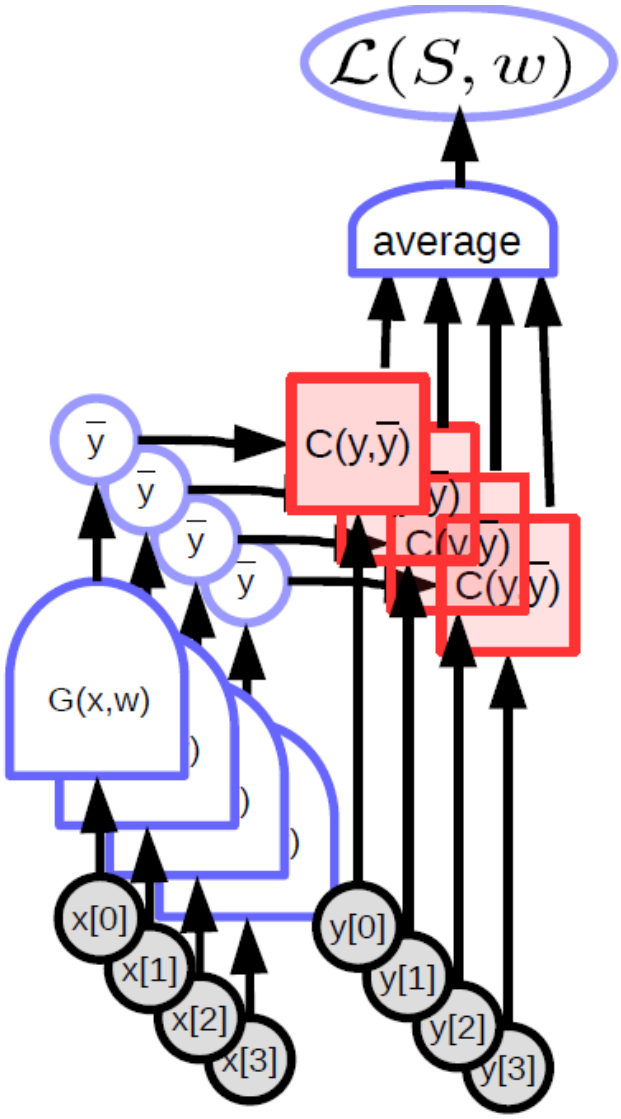
Lecture 2 Backpropagation

Created	@May 20, 2021 3:57 PM
Tags	



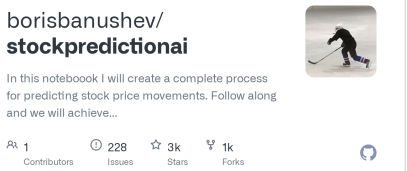
In the standard Supervised Learning paradigm, the loss (per sample) is simply the output of the cost function. Machine Learning is mostly about optimizing functions (usually minimizing them). It could also involve finding **Nash Equilibria between two functions like with GANs (check more on this papers)**. This is done using Gradient Based Methods, though not necessarily Gradient Descent.

some examples about GAN in stock trading

borisbanushev/stockpredictionai

In this noteboook I will create a complete process for predicting stock price movements. Follow along and we will achieve some pretty good results. For that purpose we will use a Generative Adversarial Network (GAN) with LSTM, a type of

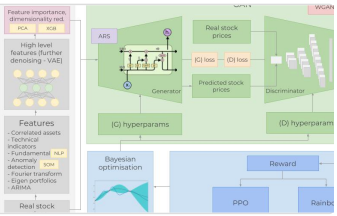
https://github.com/borisbanushev/stockpredictionai#corrassets



Using the latest advancements in deep learning to predict stock price movements

Link to the complete notebook: https://github.com/borisbanushev/stockpredictionai In this notebook I will create a complete process for predicting stock price movements. Follow along and we will achieve some pretty good results. For that purpose we will use a

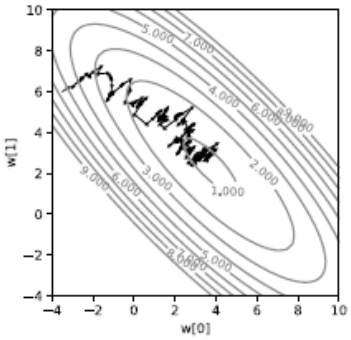
https://towardsdatascience.com/aifortrading-2edd6fac689d



RL (Reinforcement Learning) involves Gradient Estimation without the explicit form for the gradient.

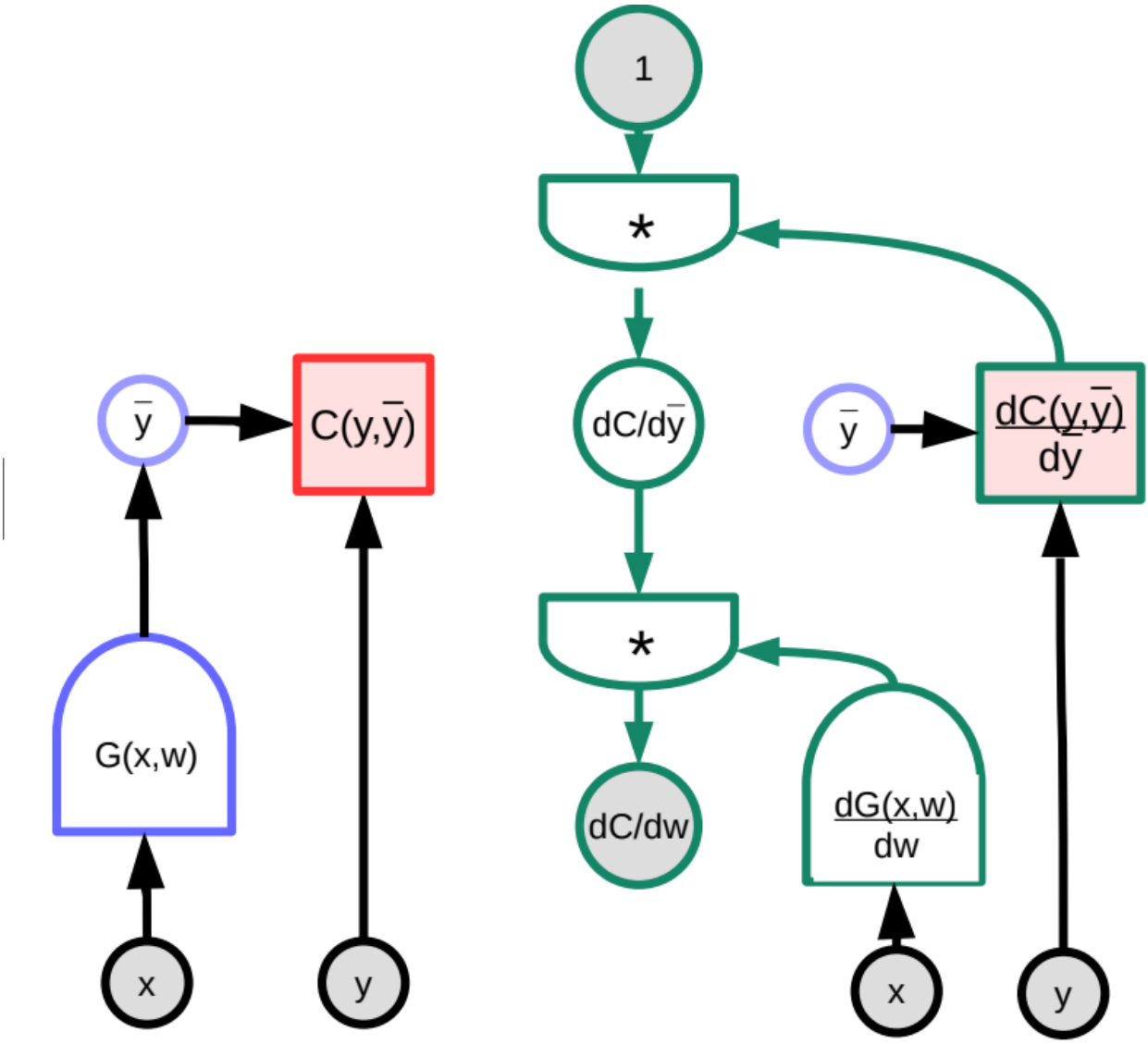
The RL cost function is not differentiable most of the time but the network that computes the output is gradient-based. This is the main difference between supervised learning and reinforcement learning. A very popular technique in RL is Actor Critic Methods. That's a way of making the cost function differentiable, or at least approximating it by a differentiable function so that one can backpropagate.

SGD



In practice, we use batches instead of doing stochastic gradient descent on a single sample.

Backpropagation



$$\frac{\partial C(y,\bar{y})}{\partial w} = 1 \cdot \frac{\partial C(y,\bar{y})}{\partial \bar{y}} \cdot \frac{\partial G(x,w)}{\partial w}$$

Practical tricks for backpropagation

1. Use ReLU as the non-linear activation function
2. Use cross entropy loss as the objective function for classification problems
3. Use stochastic gradient descent on minibatches during training
4. Shuffle the order of the training examples when using stochastic gradient descent

5. Normalize the inputs to have zero mean and unit variance (depends? There are papers showing that the raw data is better.)
6. Use a schedule to decrease the learning rate
7. Use L1 and or L2 regularization for weight decay
8. Weight initialization Kaiming initialisation(standard deviation of the weights is inversely proportional to square root of number of inputs.)
9. Use dropout (mainly in Fully Connected Neural Network, usually set 0.5 or 0.3)

深度学习中Dropout原理解析

目录： 1.1 Dropout出现的原因 1.2 什么是Dropout 2. Dropout工作流程及使用 2.1 Dropout具体工作流程 2.2 Dropout在神经网络中的使用 3. 为什么说Dropout可以解决过拟合 4. Dropout在Keras中源码分析 ...

知 https://zhuanlan.zhihu.com/p/38200980