
Project Thesis

Julie Johanne Uv

Contents

1	Introduction	2
2	Financial theory	3
2.1	Efficient Market Hypothesis	3
2.2	Arbitrage	3
2.2.1	Assets	4
2.2.2	Derivatives	5
3	Monte Carlo	7
4	Black-Scholes?	8
5	Artificial Neural Networks	9
5.1	Deep Feedforward Networks	9
5.1.1	Activation function	11
5.2	Gradient-Based Training	11
5.2.1	Backpropagation	11
5.2.2	Stochastic Gradient Descent	11

Contents

Chapter 1

Introduction

Chapter 2

Financial theory

2.1 Efficient Market Hypothesis

Efficient Market Hypothesis - why do we model stocks as a log normal random walk? Financial Market? The financial market is a common term for markets that trade different financial instruments such as stocks, bonds, currencies and different financial derivatives such as futures, forwards, options and swaps. The Efficient Market Hypothesis is stated in several ways, but essentially amounts to the properties that

- the current market price reflects history up until the present, but does not hold any further information about the future
- that the market moves immediately with the arrival of new information.

From the former it follows that stock prices has the Markov Property, i.e. the future price depends only upon the present price.

$$P(X_{t+1} = x_{t+1} | X_t = x_t, \dots, X_0 = x_0) = P(X_{t+1} = x_{t+1} | X_t = x_t) \quad (2.1)$$

From the latter it follows that the prices will move with the arrival of new information and as information essentially moves randomly, it is reasonable to model the prices randomly as well. From the efficient market hypothesis it also follows that assets are always traded at their fair value(reference) and one can never outperform the market. This can be demonstrated by the following - assume a stock is undervalued or overvalued. Then as market participants discovers, the price will immediately adjust. The Efficient Market Hypothesis does not mean that one can never profit from the market, but that one must do so at a higher risk. (Not perfectly efficient in practice.)

2.2 Arbitrage

(A result from EMH - there is no arbitrage.) Arbitrage is a risk-free instantaneous profit which can be made by exploiting price differences in the same or similar financial

instruments or assets (by simultaneous purchase and sale). Arbitrage opportunities will occur as the market is not perfectly efficient, but prices will adjust quite soon(?). Hence it is reasonable to assume market efficiency and thus no arbitrage in financial pricing models.

2.2.1 Assets

A financial asset is an asset in which its value comes from a so called contractual agreement - an agreement that cash flows will be paid to the purchaser at times specified by the contract. Examples of financial assets are cash, equities, indices and commodities. As argued, a financial asset (traded on the market) may be modeled randomly and it is common to value it's price movement by a stochastic differential equation of the following form.

$$dS(t) = u(S, t)dt + w(S, t)dW(t) \quad (2.2)$$

Here $S(t)$ is the value of the asset price at time t , u and w are arbitrary functions of asset price and time. dX is a normally distributed random variable (appendix on r.v?) with the following properties

$$E[dW(t)] = 0 \quad (2.3)$$

$$Var[dW(t)] = dt. \quad (2.4)$$

$dW(t)$ follows what is called a Wiener process(or a Brownian motion)(appendix?) and represents the randomness of the asset (without it we would just have a deterministic differential equation). (u is the drift term and w is the diffusion term.)

Stocks:

By issuing stocks to investors, a company can raise capital. If the company increases it's revenue, the value of the stocks increases. In addition the company can decide to pay out a part of the revenue as dividend per share to its shareholders.

When one values stocks, it is more reasonable to look at the relative change in price rather than the actual level of the price, i.e. $dS(t)/S(t)$ rather than $S(t)$. The contributions to the relative change also known as the return is a drift term containing the average growth of the asset price, μ , and a random term containing the volatility measuring the standard deviation of the returns, σ .

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t). \quad (2.5)$$

Again, here $dW(t)$ is the normally distributed random variable as stated above. (Another incentive/reason to evaluate the returns rather than the level is that they tend to increase exponentially by observation.)

μ and σ can both be functions of S and t . The most basic model is taking μ and σ to be constants. The unique solution to (2.5) is (Ito's Lemma appendix?)

$$S(t) = S(0) \exp((\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)) \quad (2.6)$$

from which one can obtain the result

$$\ln \left(\frac{S(t)}{S(0)} \right) \sim N \left(\left(\mu - \frac{1}{2} \sigma^2 \right) t, \sqrt{t} \right). \quad (2.7)$$

(Using this one can estimate the parameters using the historical logarithmic returns by the following.

$$r_i = \log \left(\frac{S_{i+1}}{S_i} \right) \quad (2.8)$$

$$m = \frac{1}{ndt} \sum_{i=0}^{n-1} r_i \quad (2.9)$$

$$\hat{\sigma} = \frac{1}{(n-1)dt} \sum_{i=0}^{n-1} (r_i - m)^2 \quad (2.10)$$

$$\hat{\mu} = m + \frac{1}{2} \hat{\sigma}^2 \quad (2.11)$$

(Argue more why lognormal random walk is reasonable? Says something about interest rates, how it also can be modelled as a stochastic variable, maybe not okay to assume its constant, depends on lifespan of option.)

2.2.2 Derivatives

Derivatives are instruments where the value is derived from or rely upon an underlying asset (or a group of assets) such as the financial assets stated above or interest rates. (It is a contract between two or more parties.) Examples of derivatives are options, futures, forwards and swaps. This thesis will (focus on) valuing options.

And option is a contract between two parties that gives the holder of the option the right but not the obligation to buy or sell the underlying asset at a specified time and at a specified price called the exercise (or strike) price, in the future. As stated the purchaser of the option is the holder of the option, while the seller is the writer. An option in which the holder has the right to buy the underlying is called a call option, and an option in which the holder has the right to sell the underlying is called a put option. Then there are several types of call and put options. American options for example can be exercised at any time up until expiry, while the one which will be examined(?) further in this thesis is the European option which can only be exercised at expiry. The payoff function for a European call option at expiry T is

$$C(S, T) = \max(S(T) - E, 0). \quad (2.12)$$

E is the exercise price. Similarly, the payoff function for a European put option is

$$P(S, T) = \max(E - S(T), 0). \quad (2.13)$$

What affects the value of an option? The value of an option is dependent on the underlying asset price and the time to maturity (for obvious reasons) which both will

change during the life of the option. In addition, the value of the option will be affected by the exercise price - e.g. the higher the exercise price of a call option, the higher the value of the option - and the interest rate. The interest rate affects the price through time value as the payoff will be received in the future. Lastly, the volatility will have an effect on the option value (measures the fluctuation, annualised standard deviation of the returns). Using a Assets, stocks, derivatives, arbitrage (Risk neutrality, pricing derivatives in a risk-neutral world) (Write about volatility. How much data to use for calculating historical volatility. Implied volatility.)

Chapter 3

Monte Carlo

Chapter 4

Black-Scholes?

Chapter 5

Artificial Neural Networks

5.1 Deep Feedforward Networks

Multilayer perceptrons (MLPs) are a type of deep feedforward networks. It is called feed-forward because the input \mathbf{x} floats forward through the network to finally produce an output \mathbf{y} . The goal of the network is to approximate some function f so that $\mathbf{y} = f(\mathbf{x})$. It does so by defining a mapping $\mathbf{y} = f^*(\mathbf{x}; \theta)$ for which it optimizes the parameters θ by learning/training.

(Difference between linear perceptron and multilayer non-linear activation function. Why can it approximate non-linear functions?)

The network is deep if it consists of two or more hidden layers. Hidden layers are classified by being between the input and the output layer. Each layer in the network defines an activation function w and initializes weights W_{ij} for each node and a bias $b^{(i)}$. If the number of hidden layers is n , this is called the depth of the network, the number of nodes in each hidden layer is m_i for $i = 1, \dots, n$, this is called the width of the network and $\mathbf{x} \in \mathbb{R}^d$, $W^{(1)} \in \mathbb{R}^{m_1 \times d}$, $W^{(i)} \in \mathbb{R}^{m_i \times m_{i-1}}$ for $i = 2, \dots, n$, $b^{(i)} \in \mathbb{R}$ for $i = 1, \dots, n$ then the process in vector notation would look like

$$\begin{aligned}\mathbf{h}^{(1)} &= w(b^{(1)} + W^{(1)}\mathbf{x}) \\ \mathbf{h}^{(2)} &= w(b^{(2)} + W^{(2)}\mathbf{h}^{(1)}) \\ &\vdots \\ \mathbf{h}^{(n)} &= w(b^{(n)} + W^{(n)}\mathbf{h}^{(n-1)}) \\ \mathbf{y} &= w^{(out)}(b^{(out)} + \mathbf{w}^T\mathbf{h}^{(n)})\end{aligned}$$

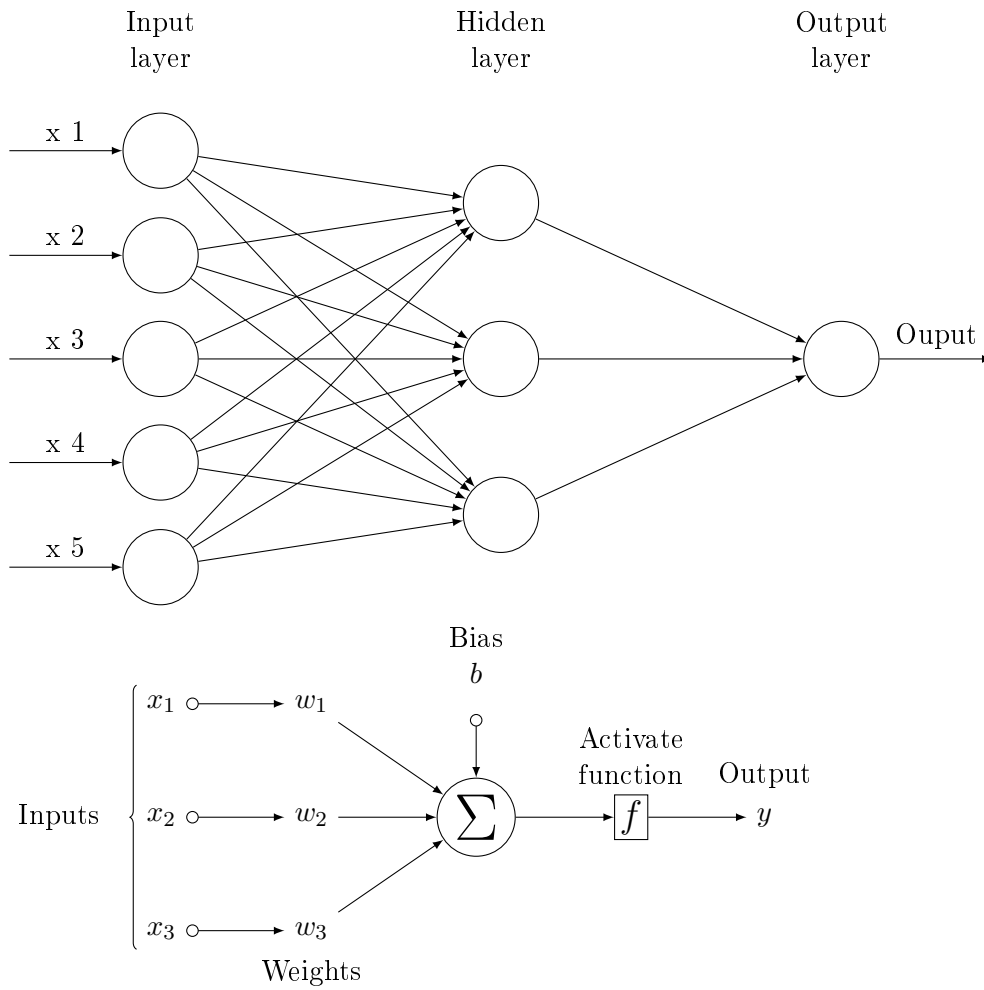


Figure 5.1: Feedforward neural network

5.1.1 Activation function

Different choices for activation functions are the sigmoid function

$$f(z) = \frac{1}{1 + e^{-z}},$$

hyperbolic tangent, \tanh ,

$$f(z) = \frac{1 + e^{-2z}}{1 - e^{-2z}},$$

rectified linear units, ReLU,

$$f(z) = \max(0, z),$$

softplus,

$$f(z) = \ln(1 + e^z)$$

and several others. (ReLU have become especially popular as of lately and is used for all convolutional networks.)

5.2 Gradient-Based Training

5.2.1 Backpropagation

In order to train the model, one needs to specify a cost function to minimise. This cost function often involves a cross-entropy between the training data and the predicted data(?) and a regularising term (penalising complex models to avoid overfitting). (write more about cost function?).

In order to minimise the cost function, the gradient of the model has to be calculated. This is done using the backpropagation algorithm.

5.2.2 Stochastic Gradient Descent

Write about the different loss functions, optimisers, activation functions, dropout rate and other regularising techniques. Argument about choice made.

(Define input and output of the neural network, why are these reasonable?)