UNIVERSIDAD NACIONAL DE CÓRDOBA

FACULTAD DE MATEMÁTICA, ASTRONOMÍA, FÍSICA Y COMPUTACIÓN.

Métodos computacionales para el cálculo de la volatilidad implícita del modelo de Black Scholes

Tesis realizada por Diego Juan Nasareno Lupi para la Licenciatura en Ciencias de la Computación en la Universidad Nacional de Córdoba

---

Dirigida por:
Dra. Noemí Patricia Kisbye
Lic. Pedro Ángel Pury

Agradecimientos

- completar
- completar
- completar

Resumen

Las finanzas cuantitativas constituyen, desde hace varias décadas, un área particular de estudio dentro de la matemática. Esta nueva disciplina surge de la necesidad de encontrar modelos cuantitativos que permitan describir el comportamiento aleatorio de activos financieros y, en particular, valorar los productos llamados derivados financieros.
Si la hipótesis sobre la dinámica de los activos es que estos siguen un proceso estocástico lognormal, con tendencia y volatilidad constante, entonces la valoración de una opción call sobre el activo está dada por la fórmula de Black Scholes. Ahora bien, dado que la volatilidad no es observable en el mercado, se define la volatilidad implícita del activo como aquella que iguala la prima del mercado con el valor dado por la fórmula.
La obtención de este parámetro de volatilidad implícita permite luego valorar otros derivados financieros como así también comprender movimientos propios del mercado.

Por otra parte, la determinación de la volatilidad implícita requiere de la aplicación de métodos numéricos, puesto que se trata de resolver una ecuación no lineal sin una solución cerrada. En los últimos años, a su vez, han aparecido propuestas de uso de métodos de aprendizaje automático para modelar de manera empírica la función que provee la volatilidad implícita.

Este trabajo incluye la exploración bibliográfica referida al concepto de volatilidad implícita y sus implicancias, y de métodos computacionales factibles de ser implementados para su cálculo. Además se realizará la implementación efectiva en computadora de algunas soluciones y se hará un análisis comparativo de la eficiencia de los distintos métodos estudiados.

Índice general

Capítulo 1

Introducción

Para el desarrollo del presente trabajo es necesario estudiar y comprender varias herramientas y técnicas que usaremos a lo largo del proyecto. A continuación presentamos un resumen de los temas que abordaremos.

1.1. Motivación

La volatilidad implícita en estos días ha adquirido una gran importancia. Se puede utilizar para el cálculo de superficies de volatilidad. Las superficies de volatilidad son usadas para valuar opciones mediante modelos matématicos. Otra utilidad es como párametro de riesgo para inversores, dependiendo de la volatilidad implícita se pondera algunas estrategias sobre otras para disminuir el riesgo y aumentar su posibilidad de rentabilidad.
Por eso es indispensable poder calcular la volatilidad implícita, generalmente se utilizan métodos iterativos para estimarla, pero estos métodos suelen ser lentos. Por lo tanto se propone el cálculo de la volatilidad implícita mediante redes neuronales para desminuir considerablemente el tiempo de cálculo.

1.2. Objetivos del trabajo

El objetivo del trabajo es determinar la volatilidad implícita sobre call europeas mediante métodos numéricos, o el uso de métodos de aprendizaje automático para modelar de manera empírica la función que provee la volatilidad implícita; ya que se trata de resolver una ecuación no lineal sin una solución cerrada.

1.3. Estructura de la Tesis

La tesis está estructurada de la siguiente manera. En el capítulo 2 se presentan los conceptos matemáticos, financieros y técnicos utilizados a lo largo del trabajo. En el capítulo 3 se muestra las implicancias de la volatilidad implícita, y la importancia de su cálculo. Los capítulos 4 y 5 refieren a la implementación de métodos para el calculo de la volatilidad implícita, tanto métodos numéricos (Bisección y Brent), como redes neuronales. En el capítulo 6 se presentan los resultados obtenidos comparando los métodos numéricos con la red nueronal. Y por último en el capítulo 7 se presenta una breve conclusión y posibles mejoras.

Capítulo 2

Nociones Preliminares

2.1. Conceptos financieros elementales

2.1.1. Los tipos de Interés

Si depositamos dinero en una cuenta bancaria, al cabo de un cierto tiempo este capital se incrementa en un determinado monto, llamado **interés**.

1. **Interés Simple**:

   $V = V_0(1+r)^t$

   Donde V representa el valor de un depósito de valor inicial V_0 transcurrido un tiempo t y r es la tasa de interés correspondiente a la unidad de tiempo utilizada, usualmente el año. ..correspondiente a la unidad de tiempo utilizada, usualmente el año.

2. **Interés Compuesto**:

   $V = V_0\left(1+\dfrac{r}{n}\right)^{nt}$

   Donde que V , V_0 y t tienen las mismas condiciones que para el interés simple, pero una tasa r de interés compuesto n veces al año.

3. **Interés Continuo**:

   El interés continuo puede verse como el interés compuesto para $n \to \infty$, es decir:

   $\lim_{n\to\infty} V_0\left(1+\dfrac{r}{n}\right)^{nt} = V_0e^{rt}$

2.1.2. Tasa libre de riesgo

A los efectos de valoración de derivados se asume la existencia de una tasa llamada **tasa libre de riesgo**. Se trata de una tasa de referencia que no tiene riesgo crediticio, es decir, que un inversor sabe que invirtiendo a esa tasa podrá recuperar el capital. Por ejemplo, los bonos del tesoro (Bonar 2025), o alguna tasa a la cual el propio estado ofrece para la devolución del préstamo (LEBAC, LETES) [1].

2.1.3. Productos básicos

Denominamos productos básicos a aquellos instrumentos financieros cuyo valor no depende de otro activo. Entre ellos están las acciones, los índices, las monedas, los commodities y los bonos.

2.1.4. Derivados

Un derivado puede ser definido como un instrumento financiero cuyo valor depende o deriva de los valores de otros activos subyacentes. Estos activos subyacentes podrían ser activos básicos u otros derivados.
En términos generales, un derivado consiste en un contrato entre dos partes para comprar o vender el subyacente. Las características y condiciones de estos contratos dan lugar a diferentes tipos de derivados, tales como contratos forward, futuros y opciones.

2.1.5. Mercados financieros

Los instrumentos financieros se comercializan en el mercado financiero. En la práctica existe un mercado formal u organizado y un mercado extrabursátil, denominado también *over the counter market* (mercados OTC) .. over the counter. En el caso del mercado formal, las negociaciones son multilaterales, es decir, existen múltiples agentes que actúan como compradores o vendedores de productos financieros. Los instrumentos que se comercializan están estandarizados, y existe una regulación de este tipo de mercados para proteger a los inversores de posibles incumplimientos de las contrapartes. En el caso del mercado extrabursátil, las negociaciones son bilaterales, es decir, entre dos partes. En estos casos los contratos suelen acordarse entre las partes en cuanto a la cantidad y características del subyacente. No existe una regulación formal sino que se basa en un principio de confianza entre partes.

El tamaño del segmento de mercados OTC es varias veces mayor que el de los mercados regulados, especialmente en los contratos de tipos de interés y de divisas. Según un estudio de ICAP[2], estima que cada día se producen 2 millones de transacciones en los mercados OTC por un nominal de 5 billones de dólares.

2.1.6. Opciones

Las opciones son contratos que dan derecho a una de sus partes a comprar (o vender) el subyacente, a un precio determinando en un tiempo futuro. Las opciones que dan derecho a compra se denominan **calls** y las que dan derecho a venta se denominan **puts**. Estos contratos tienen un valor inicial denominado prima, que es el precio al cual compra el contrato quien adquiere el derecho a compra o venta.
Quien compra una opción está en posición long sobre el contrato, y quien la vende está en posición short.
El agente que esté en posición long tiene derecho a ejercerla o no. En cambio el que este en posición short tiene una obligación sobre lo que haga su contraparte.
Las opciones que se negocian en mercados formales se denominan *opciones vanilla o estándar*.
Dentro de las opciones vanilla existen dos tipos:

- **Opciones Europeas**: son aquellas cuyo ejercicio ocurre sólo en la fecha de madurez.

- **Opciones americanas**: son aquellas que pueden ser ejercidas en cualquier momento previo a la madurez.

En el mercado OTC (.. definir acá porque es la primera vez que aparece) se negocia una variedad mucho mayor de opciones y se denominan en general *opciones exóticas* [1]. Dentro de las *opciones exóticas* se agrupan a todas aquellas opciones que tienen una mayor complejidad en comparación con las opciones vanilla. Algunas opciones exóticas son:

- **Opciones look-back**: son aquellas cuyo payoff depende del valor máximo, o del valor mínimo, que haya alcanzado el subyacente desde el inicio del contrato hasta su madurez.

- **Opciones asiáticas**: son aquellas cuyo payoff depende del promedio de valores que ha tomado el subyacente durante la vigencia del contrato, o de una parte de ese tiempo.

- **Opciones barrera**: son aquellas cuyo payoff depende de que el subyacente haya cruzado una determinada barrera a lo largo de la vigencia del contrato.

Otras opciones exóticas son las binarias, bermudas, choice, shout, basket, exchange, y muchas otras.

.. Opcion Europea no amerita una sección y lo que se dice vale para todo tipo de opción
En cada contrato se fija entonces un precio de ejercicio o **strike**, que es el precio pactado al cual se comprará o venderá el subyacente, en la fecha de expiración o también llamada **madurez** del contrato. Así, un inversor que negocia una opción adquiere una de las siguientes posiciones en el contrato:

- posición long: quien compra la opción, y por lo tanto tiene derecho a ejercerla.
- posición short: quien vende la opción o suscriptor, y por lo tanto contrae una obligación.

2.1.7. Payoff y Ganancia

El payoff de una opción europea es el valor del contrato en su madurez en función del valor del subyacente. El costo inicial o prima de la opción no se incluye en el cálculo. Si K es el precio strike y $S(T)$ es el precio final del subyacente, entonces una opción call se ejerce sólo si $S(T) > K$ ya que en caso contrario el inversor en posición long preferirá comprar el subyacente en el mercado. En el caso de una put, el inversor en posición long ejercerá la opción sólo si $K > S(T)$. Así, el valor del contrato en su madurez está dado por la ganancia o pérdida neta del inversor en caso de que se ejerza o no la opción. El Cuadro 2.1 resume los payoff de las opciones call y put europeas. Observar las Figuras 2.1 2.2.
Los diagramas de payoff para cada una de las posiciones en una opción europea con strike K son los gráficos del Payoff en función del precio final del activo subyacente. El beneficio o ganancia real del inversor incluye además el costo de la prima (Figuras 2.3 2.4).

$$\text{Payoff} \pm \text{prima}$$

donde $\pm$ dependerá de la posición long $(-)$ o la posición short $(+)$. **??**

Cuadro 2.1: Payoffs de call y put europeas

Opción	Payoff
Long en una Call	$\max(S(T) - K, 0)$
Short en una Call	$-\max(S(T) - K, 0) = \min(K - S(T), 0)$
Long en una Put	$\max(K - S(T), 0)$
Short en una Put	$-\max(K - S(T), 0) = \min(S(T) - K, 0))$

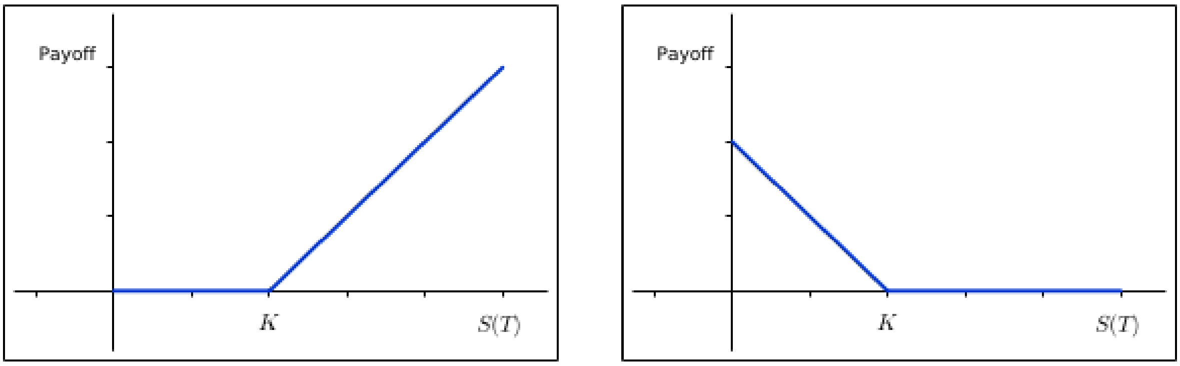

Figura 2.1: Payoff de posiciones long en una call y en una put con strike K y madurez T.

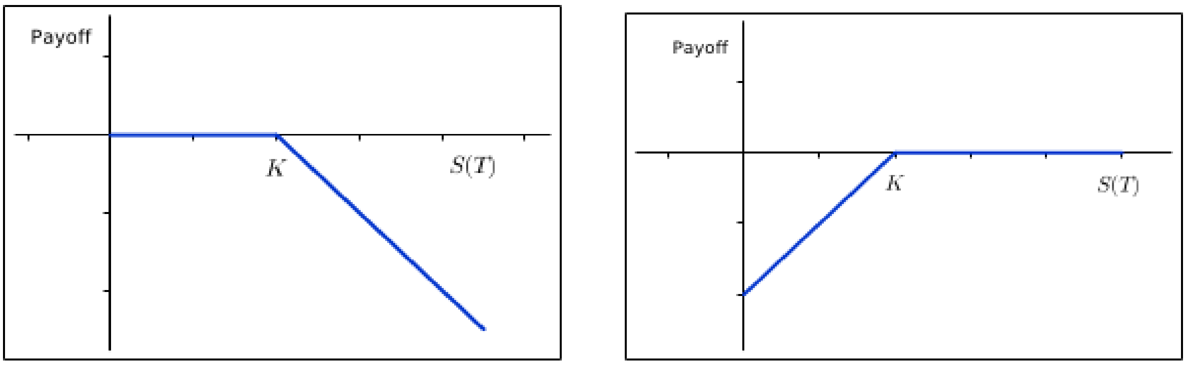

Figura 2.2: Payoff de posiciones short en una call y en una put con strike K y madurez T.

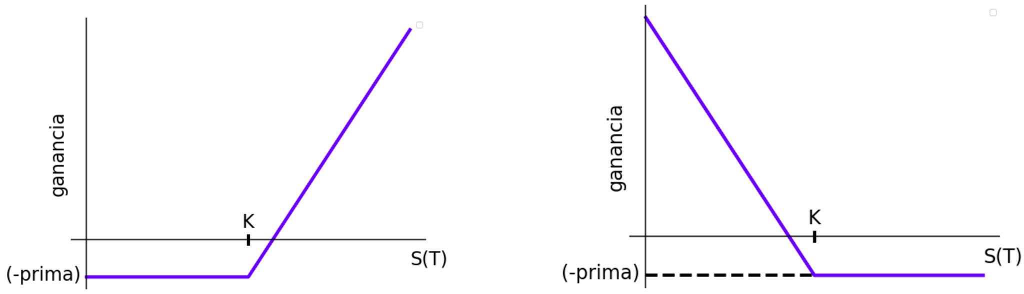

Figura 2.3: Ganancia de posiciones long en una call y en una put con strike K y madurez T.

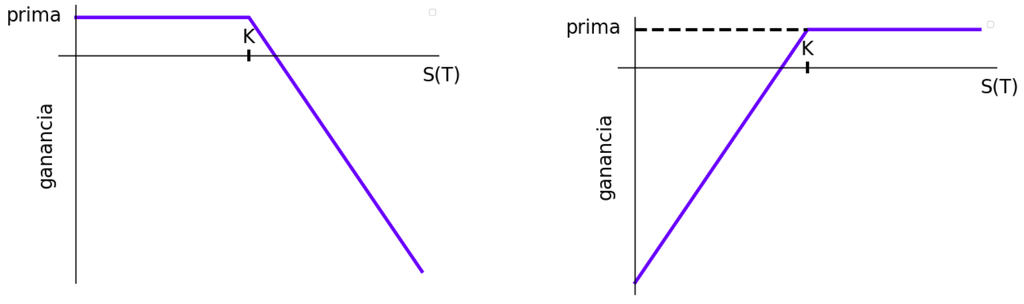

Figura 2.4: Ganancia de posiciones short en una call y en una put con strike K y madurez T.

2.1.8. Estrategias con opciones

Las estrategias son mecanismos que utilizan los inversionistas para obtener ganancia, o cubrirse de posibles casos adversos. A la convinación de diferentes opciones y activos se las llama **portfolio**.
Un **portfolio** o cartera, es un conjunto de activos, derivados e inversiones bancarias que puede tener un inversor. El valor de este portfolio es la suma de los valores de sus componentes, con un signo positivo o negativo según estén a favor o no del inversor. Las posiciones long en un activo o derivado financiero y una cuenta bancaria son valores positivos. Las posiciones short y las deudas tienen valor negativo.
Veremos algunos portfolios específicos construidos con opciones.

Spreads

Los *spreads* son estrategias que utilizan opciones del mismo tipo. Esto es: todas call o todas put. Un inversor que cree que el mercado está en alza utiliza un *bull spread*, mientras que si cree que está en baja utiliza un *bear spread*. Un *bull spread* con calls se construye con una posición long en una call, con strike K_1, y 1 posición short en una call, con strike K_2, con $K_1 < K_2$ y ambos con la misma fecha de expiración T . Luego el payoff está dado por:

Payoff(bull spread con calls) = $\max\{S(T) - K_1, 0\}$ - $\max\{S(T) - K_2, 0\}$

Notemos que el payoff es no negativo, y su valor es positivo si $S(T)$ es mayor que K_1. La ganancia se observa en la Figura 2.5.

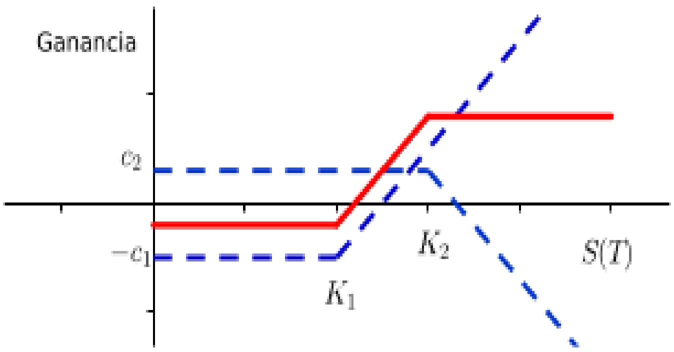

Figura 2.5: Diagrama de ganancia del bull spread con calls

Sin embargo, un *bull spread* con calls tiene un costo inicial $c_1 - c_2$. Por ello, si el valor de la acción baja habrá una pérdida, aunque acotada.

Combinaciones

Las combinaciones, son estrategias que consisten en posiciones en distintos tipos de opción, tanto call como puts. Por ejemplo, un *straddle* se construye con una posición long en una call y posición long en una put, ambos con la misma fecha de expiración T y strike K.
El payoff de una *straddle* está dado por:

Payoff(bull spread con calls) $= \max\{S(T) - K, 0\} + \max\{K - S(T), 0\} = |S(T) - K|$.

La Figura 2.6 muestra la ganancia del straddle. Notemos que el costo inicial está dado por $c + p$.

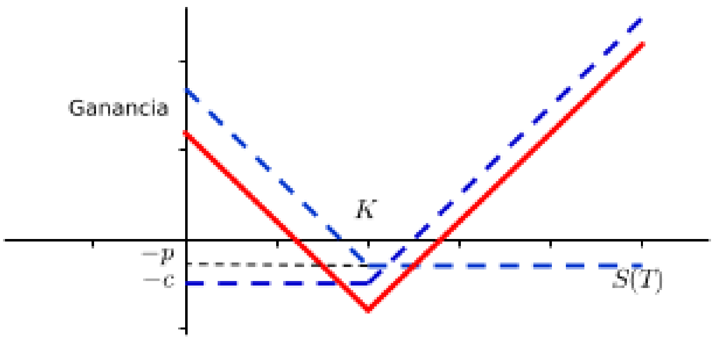

Figura 2.6: Diagrama de ganancia para un straddle

Patricia: Yo diría de hacer una sección que se llame "Modelo de Black Scholes", allí presentar la hipótesis de este modelo:

- no arbitraje (que habría que definirlo)
- completo (definirlo)
- precio de un activo que sigue un MGB
- Fórmula de BS que se obtiene a partir de las hipótesis anteriores

2.1.9. Modelo de Black Scholes

Esta fórmula comienza con el diseño de un modelo, denominado Modelo de Black-Scholes, en el cual describen el comportamiento de los precios de las

acciones, bajo ciertas hipótesis de mercado 2.1.9. A partir de esto, logran derivar una ecuación diferencial, la ecuación de Black-Scholes que es satisfecha por las opciones call y put europeas, y cuya resolución permite dar el valor exacto de la prima de la opción en un escenario de mercado sin arbitraje. La solución de esta ecuación diferencial es conocida como la Fórmula de Black-Scholes.
Los supuestos hechos por Black y Scholes cuando derivaron su fórmula de valoración de opciones fue el siguiente:

- no arbitraje: No es posible invertir en un portfolio con costo cero y que exista una probabilidad positiva de ganancia futura y una probabilidad nula de pérdida.
- completo (definirlo) Preguntar
- precio de un activo que sigue un Movimiento Geométrico Browniano [1].

2.1.10. Fórmula de Valoración de una call europea con Black-Scholes

La hipótesis del modelo de Black Scholes supone que el activo S(t) se comporta de acuerdo a un movimiento geométrico browniano, con tendencia μ y volatilidad σ. Sin embargo se demuestra que, bajo una hipótesis de no arbitrage es posible realizar un cambio de medida bajo la cual S(t) sigue un movimiento geométrico browniano con tendencia r y volatilidad σ, siendo r la tasa libre de riesgo con capitalización continua. Esta medida de probabilidad también se **llama medida de probabilidad neutral al riesgo** ..falta breve introduccion y cita al trabajo original?
..mejor no usar lenguaje tan formal como "sea", yo pondría "Consideremos que c es ..."
Consideremos que c es la prima de una opción call europea, con strike K y madurez T , sobre un activo cuyo precio sigue un movimiento geométrico browniano con tendencia $r - \sigma^2/2$ y volatilidad σ bajo las probabilidades de riesgo neutral. Sea r la tasa libre de riesgo. Entonces, bajo una hipótesis de no arbitraje se cumple que:

$$c = S(0)\Phi(d_1) - Ke^{-rT}\Phi(d_2) \tag{2.1}$$

donde:
Φ es la distribución normal estándar acumulada [3]

$$d_1 = \frac{Log\left(\frac{S(0)}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \tag{2.2}$$

$$d_2 = d_1 - \sigma\sqrt{T} \tag{2.3}$$

..itemizar, mejor redactar: donde $S(0)$ es el precio ..., r la tasa ..., ... y $sigma$...
con: $S(0)$ es precio inicial del activo, r la tasa libre de riesgo, K el strike, T la madurez de la opción y σ la volatilidad de la opción.
Para simplificar la notación vamos a utilizar S_t en vez de $S(t)$, $\forall t > 0$. .. Patricia: yo usaría S_t, $t > 0$. Ya sé que es lo mismo, pero el n suena a discreto

2.1.11. Volatilidad

La volatilidad mide la incertidumbre acerca del precio futuro de un activo. En teoría, la volatilidad se calcula continuamente, para valuar las opciones, tal como se dan los cambios en el valor de S. Black Scholes asume σ constante, esto implica una previa estimación estadística de σ, por ejemplo medir el comportamiento de los precios en los últimos meses y usar estimadores de varianza(volatilidad Histórica).
Como vimos en la sección 2.1.10 la fórmula de Black Scholes para una opción call depende de distintos parámetros, el precio inicial del activo, la tasa libre de riesgo, el stike, la madurez de la opción y su volatilidad. Pero esta ultima no te la brinda el mercado.

Volatilidad Histórica

La volatilidad histórica como su nombre lo indica, ..(Patricia: no usar primera persona singular. De usar primera persona se usa el plural, pero viendo el resto del escrito diría que uses tercera persona) muestra el riesgo histórico de un período de tiempo de hoy hacia atrás. Se calcula midiendo las variaciones que han tenido los rendimiento del activo en cierto período de tiempo, que puede ser 20 días, 50 días o 200 días o el que cada analista considere mejor y esta volatilidad por lo regular se presenta anualizada. La metodología mas común para calcularla es calculando las desviaciones estándar de los rendimientos del activo. La fórmula 2.6 muestra el cálculo de la volatilidad histórica.

$$x_t = ln\left(\frac{S_t}{S_{t-1}}\right) \tag{2.4}$$

$$\overline{X} = \frac{1}{n}\sum_{t=1}^{n} x_t \tag{2.5}$$

$$HV = \sqrt{n * \frac{1}{n-1}\sum_{t-1} n(x_t - \overline{X})^2} \tag{2.6}$$

Donde S_t es el precio del activo en el dia t, n la cantidad de dias sobre la cual calcularemos la volatilidad historica, en el caso de la volatilidad histórica anualizada $n = 252$ ya que es la cantidad de dias comerciales al año.
..Patricia: Escribiría una fórmula acá, para que además se entienda lo de anualizada

Volatilidad Implícita

La volatilidad implícita es un factor clave en el trading de opciones que refleja las expectativas respecto al desplazamiento futuro del precio de un activo subyacente. Alta volatilidad implícita indica que los participantes del mercado esperan un movimiento mayor en el futuro y, por el contrario, baja volatilidad apunta hacia un desplazamiento posiblemente menor. ..Patricia: Revisar redacción de este párrafo
Como vimos en la sección 2.1.10 para valorar una call europea necesitamos conocer, el precio inicial del subyacente, la tasa libre de riesgo, el strike, el tiempo de madurez de la opción y la volatilidad del subyacente. Los primeros cuatro parámetros son conocidos al momento de iniciar la opción. En cambio σ representa una volatilidad del activo en el período de vigencia de la opción, y por lo tanto es desconocido. Más aún, se está suponiendo constante cuando en la práctica puede ser un valor variable e incluso estocástico. Dado que las opciones call cotizan en el mercado, también es conocida la prima de la opción. Por ello se denomina volatilidad implícita al valor de σ que iguala la prima de la opción con la correspondiente fórmula de Black-Scholes.

2.1.12. Superficie de volatilidad

Para valorar las opciones los agentes usan la superficie de volatilidad, ya que la volatilidad depende del strike y de la madurez.
Una superficie de volatilidad es una representación tridimensional de las volatilidades implícitas de un subyacente en relación con los diferentes precios

Cuadro 2.2: Superficie de volatilidad

	K/S_0				
	0.9	0.95	1.00	1.05	1.10
1 mes	14.2	13.0	12.0	13.1	14.5
3 meses	14.0	13.0	12.0	13.1	14.2
6 meses	14.1	13.3	12.5	13.4	14.3
1 año	14.7	14.0	13.5	14.0	15.1
2 años	15.0	14.4	14.0	14.5	15.1
5 años	14.8	14.6	14.4	14.7	15.0

de ejercicio y las diferentes fechas de madurez. ..Patricia: madurez, o aclarar que se usará madurez o vencimiento

La superficie de volatilidad combina las sonrisas de volatilidad [4] .. Patricia: (falta definir) con la estructura temporal de la volatilidad para tabular las volatilidades adecuadas. De este modo poder valorar una opción con cualquier precio de ejercicio y fecha de vencimiento. Un ejemplo de superficie de volatilidad que puede ser usada para opciones sobre divisas como se muestra en el Cuadro 2.2. En este caso asumimos que la sonrisa es medida como la relación entre la volatilidad y $\frac{K}{S_0}$.

Una dimensión de la tabla es $\frac{K}{S_0}$ (segunda fila) y la otra es el tiempo de madurez (primera columna). Los valores del cuadro son las volatilidades implícitas calculadas a partir de la fórmula de Black Scholes. ..Patricia: ...calculadas a partir de la fórmula de Black Scholes. (Black Scholes no es un método). Y las volatilidades que no se encuentren en la tabla son calculadas mediante interpolación, por lo general se utiliza Spline cúbicos.[5] ..Patricia: por ejemplo splines cúbicos. (no sé si Spline a secas queda claro qué es)

Patricia: le falta un caption a esta tabla, y también decir que la primera columna es la madurez T. El caption diría que la tabla muestra la volatilidad sigma en la fórmula de BS de una call europea, con strike K sobre un activo con precio inicial S_0, y madurez T, bajo la hipótesis de no arbitraje. No sé quién es r acá

..Patricia: de los valores intermedios. Esto como una sección suelta no vale la pena. Yo mencionaría el teorema en el momento que se utilice, es bien básico

2.2. Método de Bisección

El método de bisección se basa en el teorema de valor intermedio ..teorema de valor intermedio para f funcion continua [6]. Si f es continua en el intervalo [a,b] y f(a)f(b)< 0, ..Patricia: usar modo matemático para f, (a,b), [a,b], , y no escribir f(a)*f(b), $sino f(a) \cdot f(b)$, o directamente f(a)f(b) entonces f tiene al menos una raiz en (a,b).
Si el algoritmo de bisección se aplica a una función continua f en un intervalo [a, b], donde $f(a)f(b) < 0$, entonces, después de n pasos se habrá calculado una raíz aproximada con un error a lo más de $\frac{(b-a)}{2^{n+1}}$ [7].

El Algoritmo 2.1 aplica el metodo de bisección.

2.3. Método de Brent

El método de Brent es un algoritmo de búsqueda de raíces que combina el método de bisección, el método secante y la interpolación cuadrática inversa. Este método converge siempre que los valores de la función sean computables dentro de una región dada que contiene una raíz.
A diferencia del método de Bisección, el método de Brent converge haciendo menos iteraciones sin perder la robustez del método de bisección ya que para ciertos casos utiliza el método de bisección, sin caer en los problemas de divergencia que tienen el metodo de interpolación cuadrática inversa, o del método de la secante.

Método de la Secante:

..Patricia: Escribir una breve introducción acá, no empezar directamente con una fórmula
En el método de la secante imita el de Newton 2.7 pero evita el cálculo de derivadas. Entonces se reemplaza $f'(x_n)$ de la fórmula de Newton por una aproximación de su derivada 2.8. [8] ..Patricia: revisar redacción

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{2.7}$$

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \approx \frac{f(x+h) - f(x)}{h} \tag{2.8}$$

Obteniendo como resultado:

$$x_{n+1} = x_n - f(x_n) \frac{x_n x_{n-1}}{f(x_n) f(x_{n-1})} \tag{2.9}$$

interpolación cuadrática inversa:

Es un metodo iterativo, que aplica la fórmula de interpolación de Lagrange para hacer una interpolación cuadrática en el inverso de f. Para hallar la raíz de la ecuación se implementa el método iterativo [9]: .. la raíz de la ecuación

$$\begin{aligned} x_{n+1} = \; & \frac{f_{n-1}f_n}{(f_{n-2}-f_{n-1})(f_{n-2}-f_n)}x_{n-2} \\ & + \frac{f_n f_{n-2}}{(f_{n-1}-f_n)(f_{n-1}-f_{n-2})}x_{n-1} \\ & + \frac{f_{n-1}f_{n-2}}{(f_n-f_{n-1})(f_n-f_{n-2})}x_n \end{aligned}$$

Donde $f_k = f(x_k)$. El algoritmo 2.2 aplica el metodo de Brent.

Algorithm 2.1: Bisección

Input : f, a, b, tol
Output: c
Precondition: f(a)f(b) $<$ 0 **and** a $<$ b

```
err := b - a;
while tol < err do
    c := (a+b)/2;
    err := err/2;
    fc := f(c);
    if fc = 0 then
        return;
    end
    if fcf(a) < 0 then
        b := c;
    else
        a := c;
    end
end
return;
```

2.4. Redes neuronales

Patricia: Introducir más el tema acá(NO ENTIENDO). El lector no tiene por qué saber qué es una red neuronal
..La red neuronal se entrena para estimar

Algorithm 2.2: Brent

Input : f, a, b, tol
Output: b
Precondition: f(a)f(b) $<$ 0

if $|f(a)| < |f(b)|$ **then**
  swap(a,b);
end
c := a;
mflag := True;
while $tol < |b - a|$ ***or*** $f(b) = 0$ **do**
  if $f(a) \neq f(c)$ *and* $f(b) \neq f(c)$ **then**
    A := $\dfrac{af(b)f(c)}{(f(a)-f(b))(f(a)-f(c))}$;
    B := $\dfrac{bf(a)f(c)}{(f(b)-f(a))(f(b)-f(c))}$;
    C := $\dfrac{cf(a)f(b)}{(f(c)-f(a))(f(c)-f(b))}$;
    s := A+B+C; (interpolacion cuadrática inversa)
  else
    s := $b - f(b)\dfrac{b-a}{f(b)-f(a)}$; (metodo de secante)
  end
  if ***not*** $\left(\dfrac{3a+b}{4} < s < b\right)$ ***or*** *(mflag* ***and*** $|s - b| \geq |b - c|/2$*)* ***or*** *(**not** mflag* ***and*** $|s - b| \geq |c - d|/2$*)* ***or*** *(mflag* ***and*** $|b - c| < tol$*)* ***or*** *(**not** mflag* ***and*** $|c - d| < tol$*)* **then**
    mflag := True;
    s := $\dfrac{a+b}{2}$; (metodo de bisección)
  else
    mflag := False;
  end
  d := c;
  c := b;
  if $f(a)f(s) < 0$ **then**
    b := s
  else
    a := s
  end
  if $|f(a)| < |f(b)|$ **then**
    swap(a,b);
  end
end
return;

La red neuronal se entrena para estimar los parámetros óptimos para una función $f(x, W, b)$, a partir de un conjunto de entrenamiento $\{(x_1, y_1), ..., (x_n, y_n)\}$, tal que $f(x_i, W, b) \approx y_i$, para $i = 1, ..., n$.

2.4.1. Perceptrón

La unidad básica de una red neuronal son los perceptrones (o neuronas). Un perceptrón toma un vector de entradas de valor real, calcula la combinación lineal de estas entradas con respecto a los pesos, luego genera una salida, dependiendo de la función de activación. Más precisamente:

$$o(x_1, ..., x_n) = \psi(w_0 + x_1 w_1 + ... + x_n w_n)$$

donde cada w_i es una constante de valor real, o peso, y ψ es la función de activación [10], como muestra la figura 2.7:

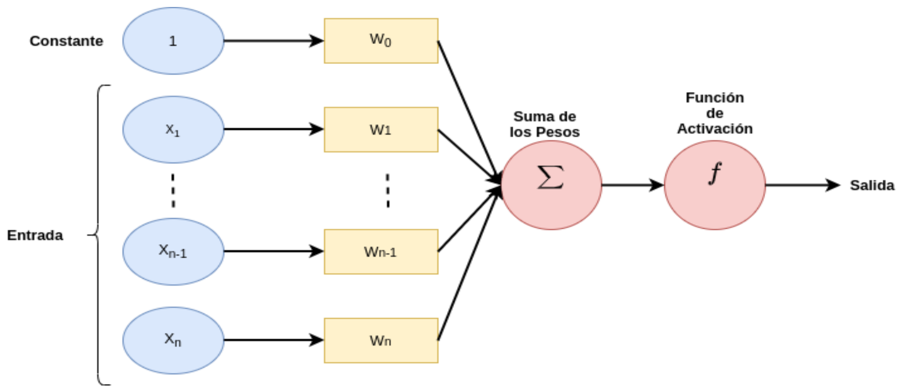

Figura 2.7: Esquema del perceptrón

2.4.2. Redes Feed-Forward

Este tipo de algoritmo recibe el nombre de "red" porque se construye componiendo funciones (perceptrones).
La arquitectura de este tipo de modelos se puede dividir en tres partes principales, denominadas capas, tal como se muestra en la Figura 2.8. ..La característica distintiva de este tipo de red es que no existen conecciones entre neuronas de una misma capa. La característica distintiva de este tipo de red es que no existen conecciones entre neuronas de una misma capa. Las capas a su vez etán completamente conectadas (o fully connected) debido a que cada neurona de una capa se conecta con todas las neuronas en la capa siguiente.

En la Figura 2.8 podemos observar las distintas capas que componen una red neuronal. En primer lugar tenemos la capa de entrada, donde se define cuántos valores de entrada tomará nuestro modelo, estos serán luego enviados a la primer capa oculta. Después de esto puede venir una o más capas ocultas, seguidas de una capa de salida que genera ..que genera una aproximación del resultado deseado, por ende, si una red neuronal Feed-Forward tiene N capas, entonces tiene N-2 capas ocultas. Cada unidad (o input) alimenta solo las unidades de la siguiente capa. Las unidades en intermedio las capas a menudo se llaman capas ocultas (hidden units) porque no tienen conexión directa con el datos, tanto de entrada como de entrada como de salida [11].

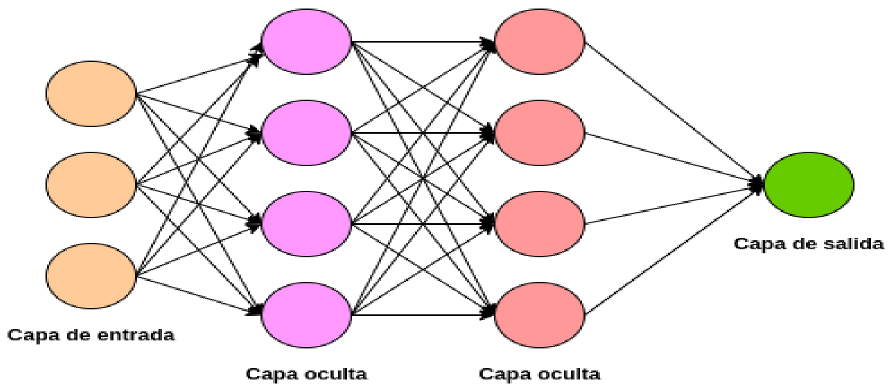

Figura 2.8: Feed-Forward

2.4.3. Aprendizaje por Descenso por el Gradiente

Como hemos mencionado anteriormente el objetivo de una red neuronal es estimar los parámetros óptimos para una función $f(x, W, b)$. Una forma de estimar esos parámetros es reducir el error medido por una función de costo ..medido por una función de costo:

$$J(x, W, b, Y) = \frac{1}{n}\sum_{i=1}^{n}(f(x_i, W, b) - y_i)^2 \tag{2.10}$$

Para reducir dicho error se busca W tal que minimize la funcion $J(x, W, b, Y)$, se puede utilizar el descenso del gradiente.
Sea $w = [W, b]$, (para simplificar la notación). El gradiente de J en w, denotado como $\nabla J(w)$, es el vector de derivadas parciales de J, es decir, $\nabla J(w) = \left(\frac{\partial J(w)}{\partial w_1}, \frac{\partial J(w)}{\partial w_2}, ..., \frac{\partial J(w)}{\partial w_n}\right)$

El descenso del gradiente es un algoritmo iterativo, se comienza ..se comienza con valor inicial de w^0 (por ejemplo $w^0 = [0, 0, ..., 0]$) y entonces en cada iteración damos un paso en la dirección negativa del gradiente en el punto actual. Esto es, $w^{t+1} = w^t - \eta \nabla J(w^t)$, donde $\eta > 0$, conocido como el learning rate, es el encargado de decidir el tamaño del paso que damos. Ya que $\nabla J(x, w^t, Y, b) > 0$ cuando J es creciente en w^t, y $\nabla J(x, w^t, Y, b) < 0$ cuando J es decreciente en w^t (como muestra la imagen 2.4), se obtiene $J(x, w^{t+1}, b, Y) \leq J(x, w^t, b, Y)$, siempre y cuando η no sea muy grande(Observar Figura 2.9).
En Algoritmo 2.3 aplica el método del descenso por el gradiente.

Algorithm 2.3: Descenso por el gradiente

Input : x, Y, η, T
Output: w
Precondition: d = len(w)

initialize w^0;
t := 0;
while $t < T$ **do**
    i := 0;
    while $i < d$ **do**
        $w_i^{t+1} = w_i^t - \eta \nabla J(x, w_i^t, Y, b)$;
        i := i+1;
    end
    t := t+1;
end
return;

..Representación esquemática del descenso por el gradiente

2.4.4. Funciones de Activación

..Esta sección cita la siguiete y no se entiende. Mejor
La función de activación en las redes neuronales son las encargadas de enviar señales al siguiente nivel o siguiente capa. Es crucial elegir la función de activacion correcta para no caer en el *problema de desvanecimiento de gradiente*.
Otra importante caractirística que debe tener la función de activación es ser diferenciable. Ya que al aplicar el algoritmo del *descenso del gradiente*, cuando se propaga para hacia atras, calcula los gradientes de error con respecto a los pesos, para calcular los nuevos pesos acordemente. [12]
Las fórmulas 2.11, 2.12, 2.13 muestra las funciones de activación que vamos

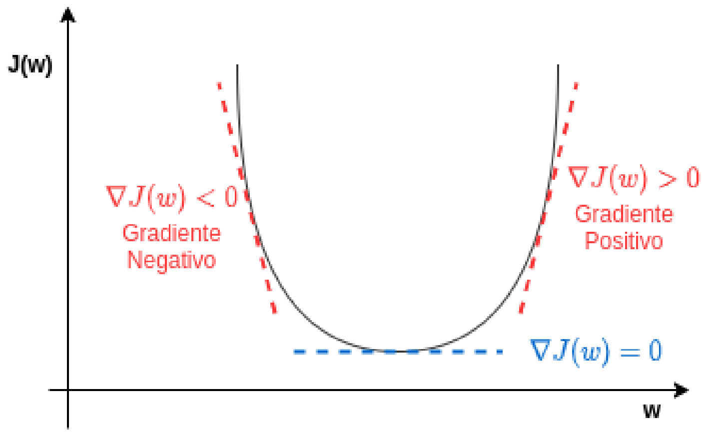

Figura 2.9: Representación esquemática del descenso por el gradiente

a utilizar, en nuestro caso el α de la Elu es 1.

$$ReLu(z) = max(z, 0) \tag{2.11}$$

$$Tahn(z) = \frac{e^z - e^-z}{e^z + e^-z} \tag{2.12}$$

$$Elu(z) = \begin{cases} \alpha(e^z - 1) & z < 0 \\ z & z > 0 \end{cases} \tag{2.13}$$

..Esto no correspponde como figura. Pasar las ecuaciones en latex y graficar esquemáticamente las funciones en una figura comparativa

2.4.5. Learning Rate

El learning rate determina el tamaño del paso en cada iteración mientras se mueve hacia un mínimo de una función de error. Un learning rate grande conduce rápidamente al un mínimo local, pero con riesgo de diverger. En cambio un learning rate pequeño no diverge pero necesita muchas iteraciones para llegar al mínimo local (Figura 2.10). Comúnmente se utiliza un learning rate grande y se lo va decrementando por cada época (es cuando un conjunto de datos completo pasa a travez de la red hacia adelante y hacia atras una vez) ..falta definir hasta encontrar un buen resultado.
..Efecto de la tasa de aprendisaje sobre el entrenamiento

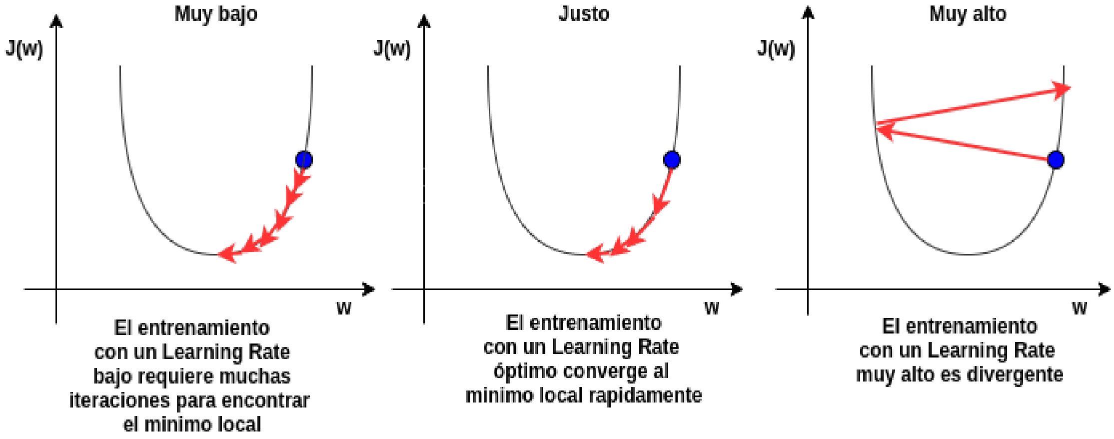

Figura 2.10: Efecto de la tasa de aprendisaje sobre el entrenamiento

Algoritmos de decrecimiento del Learning Rate:

Time-Based Decay: $base\_lr \dfrac{1}{1 + decay * epoch}$ [13]

Step Decay: $base\_lr * dacay^{floor\left(\dfrac{1+epoch}{epoch\_drop}\right)}$ [13]

Exponential Decay: $base\_lr * e^{(-k*epoch)}$ [14]

Donde epoch representa la época en cual la red se encuentra, base_lr es el valor inicial del learning rate. Con decay y epoch_drop como variables. ..Falta definir varias cantidades

Cyclical Learning Rate:

En este caso el learning rate varía entre un mínimo y máximo, creciendo y decreciendo como muestra la Figura 2.11. Para obtener dicho máximo y mínimo utilizaremos un método similiar al de Smith [15] para determinar el learning rate. A diferencia del método de Smith, iremos subiendo el learning rate exponencialmente por cada iteracion de la red. ..falta definir. Empezaremos con un learning rate de 10^{-10} hasta llegar a 1, comparando la tasa de error contra el learning rate.

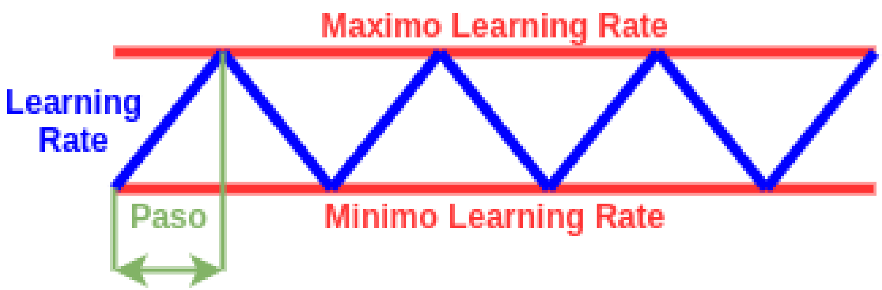

Figura 2.11: Cycical Learnig Rate

2.5. K-fold cross validation

K-fold cross validation es una técnica utilizada para evaluar modelos propuestos, con el fin de encontrar el mejor modelo. En K-fold cross validation los datos de prueba se dividen en K subconjuntos. Uno de los subconjuntos se utiliza como datos de test y el resto (K-1) como datos de entrenamiento. El proceso de cross validation es repetido durante K iteraciones, con cada uno de los posibles subconjuntos de datos de test. Finalmente se realiza el promedio de los resultados de cada modelo(cada modelo tiene K resultados), guardando el modelo que obtuvo mejor promedio. Este método es muy preciso puesto que evaluamos a partir de K combinaciones de datos de entrenamiento y de test.

Capítulo 3

Implicancias de la volatilidad implícita

En esta sección describiremos diferentes usos de la volatilidad implícita.
El primer uso es en el cálculo de superficies de volatilidad, la cual sirve para valorar otros tipos de opciones.
El segundo uso es utilizar la volatilidad implícita como indicador de riesgo de mercado, de esta manera poder ejercer estrategias, ya sean de cobertura o de especulación.

3.1. Superficie de Volatilidad para valorar opciones

Para valorar las opciones europeas, lookback, asiáticas, entre otras, existen modelos matemáticos para calcular Capítulo anterior..no están explicadas las opciones exóticas el valor la opción. A continuación listaremos una serie de modelos desarrollados para valorar distintos tipos de opciones vainilla y exóticas. [16]

- Black-Scholes(1973): Modelo para valorar opciones europeas.
- Rubinstein (1991): Modelo para valorar opciones chooser simples.
- Conze y Viswanathan(1991): Modelo para valorar opciones lookbacks con precio de ejercicio fijo.
- Goldman, Sosin y Gatto(1979): Modelo para valorar opciones lookbacks con precio de ejercicio flotante.

- Kemma y Vorst (1990): Modelo para valorar opciones asiáticas con media geométrica.
- Levy (1992): Modelo para valorar opciones asiáticas con media aritmética.
- Margrabe (1978): Modelo para valorar opciones sobre el intercambio de dos activos.

Todos los modelos anteriormente mencionados para ser aplicados necesitan la volatilidad entre otros parámetros. Pero si el mercado no nos brinda el valor de la volatilidad, una manera de estimarla es mediante la superficie de volatilidad, definida en el capítulo anterior. Capítulo anterior..Revisar esto. Para definir la superficie de volatilidad se necesita del cálculo de la volatilidad implícita. Entonces no se entiende cómo es que se estima la volatilidad a partir de la superficie de volatilidad.

3.2. Estrategias dependiendo de la volatilidad implícita

Entender el concepto de volatilidad es esencial para tener éxito en la comercialización de opciones. Un inversor que puede reconocer cuándo una opción o series de opciones están baratas o caras tiene una gran ventaja a la hora de invertir. ..Esto quiere decir (no se entiende la implicancia) Esto quiere decir que si la volatilidad implícita es baja, entonces el prima de la opción va a ser baja. En cambio si la volatilidad implícita es alta, el prima de la opción va a ser alta. Ya que el prima de la opción es creciente con respecto a la volatilidad. [1] ..¿Esto tiene que ver con que la prima de una call(la put tambien por paridad de put-call) es una función creciente con respecto a la volatilidad?)
Ahora bien, ¿cómo reconocer cuándo la volatilidad implícita es alta o baja?. Una manera es comparar la volatilidad histórica de los ultimos 2 años con la volatilidad implícita de la opción. ..en todo caso se calcula la volatilidad histórica en dos años, no se entiende lo de la 'volatilidad diaria'. Por ejemplo si la volatilidad implícita es considerablemente menor a su volatilidad histórica de los últimos 2 años, entonces la volatilidad implícita es baja.
Ahora nombraremos algunas estrategias en las cuales se hace uso de la volatilidad implícita.[17]

Terminología que usaremos en esta sección:

- ATM(at-the-money): Una opción está ATM cuando $S(0) = K$ tanto en opciones put como call.
- OTM(out-of-the-money): Una opción call está OTM cuando $S(0) < K$ y una opción put cuando $S(0) > K$.
- ITM(at-the-money): Una opción call está ITM cuando $S(0) > K$ y en una opción put cuando $S(0) < K$.

Donde $S(0)$ es el precio del subyacente al momento de entrar a la opción y K es el strike de la opción.

3.2.1. Backspread

Las claves para aplicar la estrategia Backspread son las siguientes:

- El inversor espera un movimiento particular del mercado, pero teniendo un poco de protección en caso de equivocarse.
- Busca volatilidad implícita baja al momento de aplicar la estrategia, con esperanza que suba en el futuro.

La estrategia Backspread involucra suscribir opciones call o put, ATM o ITM (primas caras), ..No has definido qué significa ATM, ITM y OTM, (ni por qué esto que decís implica primas caras.) y simultaneamente comprar un número mayor de opciones OTM (primas baratas), por principio de no arbitraje. Idealmente se busca obtener que la prima de las opciones suscritas sea mayor al valor de las opciones compradas.
..prima? strike?)
En el caso que tengamos expectativas de un movimiento alcista del subyacente entonces la estrategía será Call Backspread, que consiste en vender y comprar opciones call. El objetivo es que el precio del subyacente suba o baje drásticamente para obtener ganancia, con preferencia alcista como se puede observar en la Figura 3.1. ..ponerle label a las figuras y citarlas con ref
En el caso que tengamos expectativas de un movimiento bajista del subyacente entonces la estrategia será Put Backspread, que consite en vender y comprar opciones put. El objetivo es el mismo que el Call Backspread, con preferencia bajista como se puede observar en la Figura 3.2. ..(usar label y ref)
..me resulta confusa la explicación de la estratrgia. Poner bien en claro la relación entre las primas y los strikes de las opciones suscritas (posición short) y compradas (posición long). Escribir las ecuaciones para la ganancia en cada caso

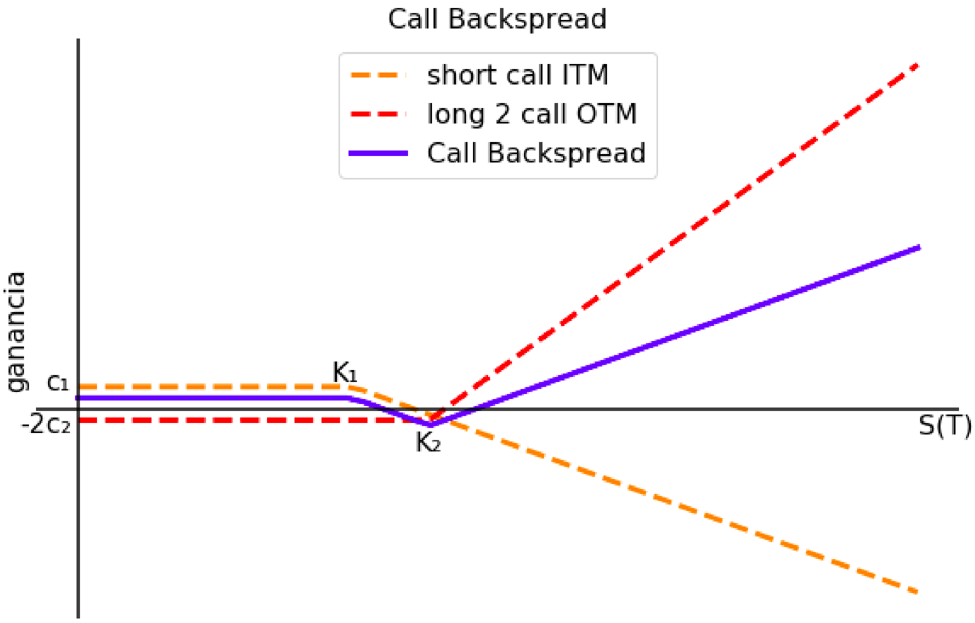

Figura 3.1: Call Backspread.

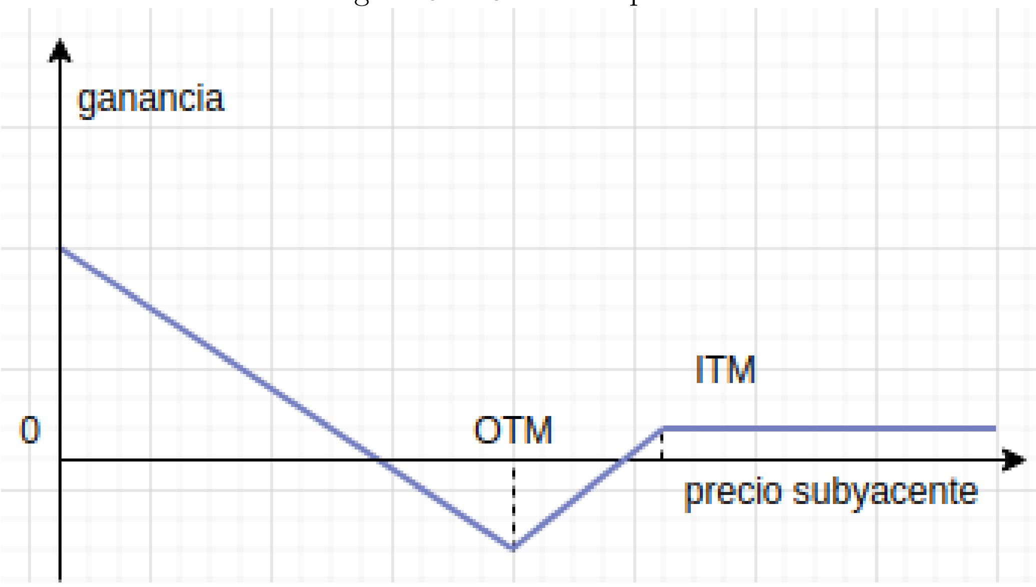

Figura 3.2: Put Backspread.

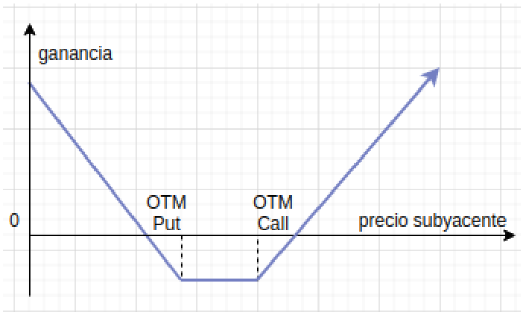

Figura 3.3: Long Strangle.

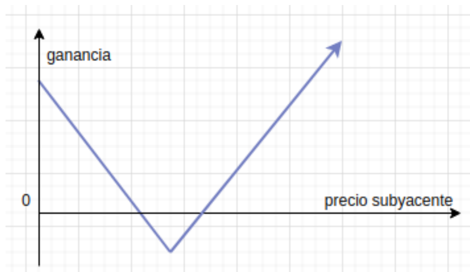

Figura 3.4: Long Straddle.

3.2.2. Long Straddle y Long Strangle

Las claves para aplicar estas estrategias son las siguientes:

- Se espera a que el valor del subyacente tenga un gran cambio, sin importar si el precio sube o baja.
- Se busca volatilidad implícita baja al momento de aplicar la estrategia.
- Idealmente se busca opciones out-of-the-money. ..Ponerle verbo a esta oración
- Se busca tiempo adecuado antes de la expiración, tratar que no sean contratos cortos. ..reescribir esta oración con un verbo

La estrategia de comprar una Straddle o una Strangle consiste en comprar una opción call y una opción put al mismo tiempo, con el mismo tiempo de madurez, ambas con el igual strike (Straddle) o diferente strike (Strangle). Ideamente se busca que al momento de comprar las opciones, la volatilidad implícita sea baja (comprar opciones baratas), y que el precio del subyacente tenga un abrupto cambio (Guerra comercial entre Estados Unidos y China, Pandemia, Precio Petróleo, entre otros) sin importar su dirección, puede ser alcista o bajista. Por eso el tiempo de madurez es clave en este tipo de estrategias. La pérdida en este tipo de estrategias es limitada, es el valor de las primas. En cambio la ganancia puede ser ilimitada.
La Figura 3.3 usar label y ref para las figuras muestra la ganancia de la estrategia de Strangle como función del precio del subyacente, mientras que la Figura 3.4 la muestra para la estrategia de Straddle.
..Los gráficos de payoff y de ganancia deberían estar explicados en los capítulos anteriores

3.2.3. Long Butterfly Spread

Las claves para aplicar la estrategia son los siguientes:

- Se espera que el valor del subyacente no varíe mucho hasta su vencimiento. ..¿en qué período?.
- Se busca volatilidad implícita alta al momento de aplicar la estrategia, cuánto más alta mejor.
- Se invierte en opciones cortas, con un tiempo de expiración menor a 60 dias.

La estrategia Long Butterfly Spread puede utilizarse tanto con opciones call, como con opciones put.
En caso que se utilicen opciones call, la estrategia consiste en comprar una opción call con un determinado strike K_1, vender dos call con un strike mayor al strike de la opción nombrada anteriormente,K_2, y comprar otra opción call con un strike mayor a todos los anteriores, K_3. Esto es, $K_1 < K_2 < K_3$.
En caso que nuestra estrategia utilice opciones put, la estrategia consiste en comprar una opción put a un determinado strike, vender dos put con un strike menor al strike de la opción nombrada anteriormente, y comprar otra opción put con un strike menor a todos los anteriores.
En esta estrategia la mayor ganancia se obtiene cuando el precio del subyacente se encuentre próximo al strike del medio(opciones vendidas). Pero al ser la volatilidad alta, el precio del subyacente es muy variable, siendo así esta estrategia riesgosa. Una buena medida es abandonar la estrategia cuando se haya obtenido una buena ganancia, sin esperar que el precio del subyacente llegue al pico, como se observa en la Figura 3.5.
..El gráfico muestra un $K_2 = (K_1 + K_3)/2$. Sería importante marcar estos valores de primas y strikes en los gráficos

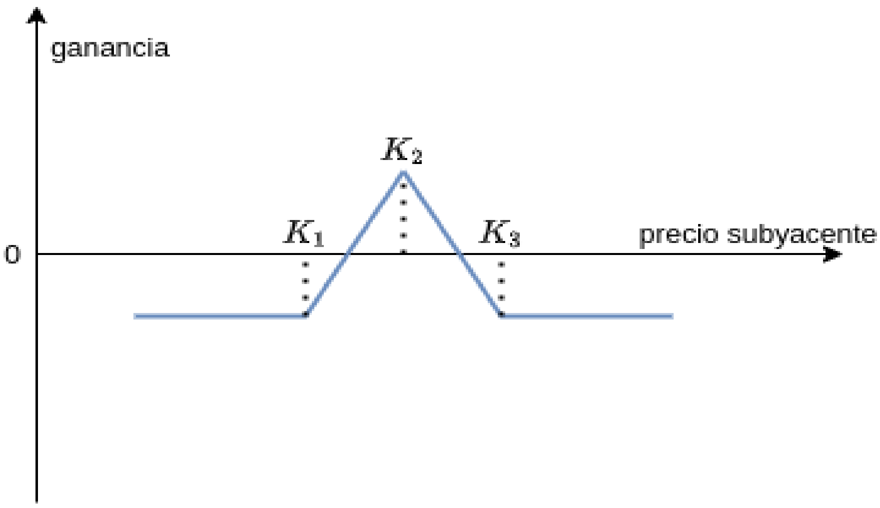

Figura 3.5: Long Butterfly Spread

Capítulo 4

Implementación númerica del método de Bisección y de Brent

En esta sección comentaremos detalles de la implementación del método de Bisección y del método Brent para aproximar la volatilidad implícita a partir de la fórmula de Black Scholes.

4.1. Introducción

Tanto el método de Bisección como el método de Brent son algoritmos de búsqueda de raíces de funciones, como lo es nuestro problema de encontrar el valor de la volatilidad implícita a partir de la fórmula de Black Scholes.
Sea c la prima de una opción call europea con madurez T y strike K sobre una acción cuyo precio actual es $S(0)$, siendo r la tasa de interés libre de riesgo. La función a la cual aplicaremos ambos métodos numéricos está dada por la fórmula 4.1.

$$g(\hat{\sigma}) = S(0)\,\Phi(d_1(\hat{\sigma})) - K\,e^{-rT}\,\Phi(d_2(\hat{\sigma})) - c \tag{4.1}$$

donde d_1 es la función 2.2, d_2 es la función 2.3, Φ es la distribución normal estándar acumulada [3] y c es la prima de una opción call europea.
..donde d_1, d_2 y Φ están dadas por las fórmulas . . . usamos la notación definida en la Sec. .. referenciafalta.
..Todo este párrafo ya fue escrito en la sección 2.1.10. c faltó definir en ese momento. Hay que numerar todas las ecuaciones en 2.1.10 para poder citarlas acá sin repeticiones
Podemos ver que g es monótona creciente con respecto a σ. Para esto si definimos $B(\sigma) = S(0)\Phi(d_1(\sigma)) - Ke^{-rT}\Phi(d_2(\sigma))$, d_1 es la función 2.2 y d_2

es la función 2.3. En la referencia [1] se demuestra que la derivada de $B(\sigma)$, $B'(\sigma) = Ke^{-rT}\Phi'(d_2(\sigma))\sqrt{T} > 0$ para todo $\sigma > 0$.
Luego como B es una función monótona creciente con respecto σ se tiene que

$$g(\sigma) = B(\sigma) - c \tag{4.2}$$

tambien lo es, ya que c es constante.

4.2. Implementación de los métodos numéricos

4.2.1. Intervalo inicial para el método de bisección y de Brent

Como mostramos de la sección anterior, $g : (0, \infty) \to \mathbb{R}$ es una función monótona creciente. Podemos tomar $a = \epsilon$, con $\epsilon \to 0$, como sabemos que la volatilidad implícita $(\hat{\sigma})$ es mayor a 0, entonces $a < \hat{\sigma}$, por lo tanto $g(a) < 0$, ya que $g(\hat{\sigma}) = 0$. ..Esto el lector no lo deduce fácilmente
Ahora para elegir b utilizaremos el siguiente algoritmo:

Algorithm 4.1: Encontrar b

Input : g
Output: b

```
b := 1;
while g(b) < 0 do
    b := b*10;
end
return;
```

4.2.2. Aplicación del Método de Bisección

Inicializamos el intervalo $[a, b]$ para aplicar el método de bisección, de manera que $a < b$ y $g(a) < 0 < g(b)$. Luego aplicamos el algoritmo de bisección 2.1 hasta hallar un ξ, tal que $g(\xi) = 0$ ó $|b-a| < \varepsilon$. A ε lo llamaremos tolerancia. Cuando esté corregido el capítulo 2 veré si esto queda claro. Yo creo que uno debe decir que va construyendo intervalos I_j ..., pero vemos después.

4.3. Aplicación del Método de Brent

Inicializamos el intervalo $[a, b]$ para aplicar el método de Brent, con $a < b$ y $g(a) < 0 < g(b)$. Luego aplicamos el algoritmo de Brent 2.2 hasta hallar un ξ, tal que $g(\xi) = 0$ ó $|b - a| < \varepsilon$. Nuevamente, a ε lo llamaremos tolerancia.

Capítulo 5

Implementación con Red Neuronal Feed-Forward

En este capítulo presentamos el cálculo de la volatilidad implícita mediante el diseño y uso de redes neuronales.
.. Esto va en el capítulo sguiente: Todo cálculo realizado en este cápitulo es sobre GPU(GeForce GTX 1080/PCIe/SSE2), RAM(15.6 GiB).

5.1. Cálculo de prima de opción call usando S/K

Rene Garcia y Ramazan Gençay [20] invocando la homogeneidad de la fórmuma de Black Scholes, demostraron que las redes neuronales estiman mejor el precio de una opción usando el cociente S/K. Así podelos reescribir la ecuación 5.1.

$$\frac{c}{K} = \frac{S(0)}{K}\Phi(d_1) - e^{-rT}\Phi(d_2) = B'(S(0)/K, T, r, \sigma) \tag{5.1}$$

donde d_1 es la ecuación 2.2 y d_2 es la ecuación 2.3.

5.2. Generación de muestra

..Falta una breve introducción a qué y para qué es la muetra
Como hemos visto en la sección anterior, el nuevo cálculo de de prima a través de la ecuación 5.1 nos dará un valor para $\frac{c}{K}$. En nuestro caso buscamos estimar la volatilidad implícita, pero usaremos el cociente para generar la muestra ya que en la práctica la red estima mejor.

Luego vamos a generar 2 muestras, una muestra amplia y una muestra estrecha. La muestra amplia va a ser para el entrenamiento de la red, y la muestra angosta para la evaluación de la red. ..Por qué? Para qué
Las variables de la muestra ya sea amplia o estrecha, pertenecerán a un ambiente previamente definido, pero las variables de la muestra estrecha sera un subconjunto de la amplia, ya que la red estima peor en los extremos. como puede verse en la Cuadro 5.2.

	Parametros	muestra amplia	muestra estrecha
Entrada	precio ratio(S_0/K)	$[A^1,B^1]$	$[A^1+C_1^1,B^1-C_2^1]$
	Tiempo de madurez(τ)	$[A^2,B^2]$	$[A^2+C_1^2,B^2-C_2^2]$
	volatilidad(σ)	$[A^3,B^3]$	$[A^3+C_1^3,B^3-C_2^3]$
	Tasa libre de riesgo(r)	$[A^4,B^4]$	$[A^4+C_1^4,B^4-C_2^4]$
Salida	Precio de Call(c/K)	(O^1,L^1)	(O^2,L^2)

Cuadro 5.1: Rango de los parámetros para la generación de las muestras $A^i+C_1^i < B^i - C_2^i, \quad C_j^i > 0, \quad O^j < L^j$, para $i = 1,2,3,4,\ j = 1,2$

Luego para generar una muestra de tamaño N, aplicaremos N veces el algoritmo 5.1, donde:

- **número_aleatorio** Es una función que dado un rango genera un número aleatorio perteneciente a dicho rango.
- **B′** Es la fórmula de Black-Scholes dada en la sección anterior.
- **rango_ratio** Es el rango de valores de el ratio(S_0/K).
- **rango_T** es el rango de valores de el tiempo de maturez(τ).
- **rango_**σ es el rango de valores de la volatilidad implícita (σ).
- **rango_**r es el rango de valores de la tasa libre riesgo(r).
- C es c/K

Algorithm 5.1: Generación muestra

Input : rango_ratio, rango_T, rango_σ, rango_r
Output: C

ratio := numero_aleatorio(rango_ratio);
T := numero_aleatorio(rango_T);
σ := numero_aleatorio(rango_ratio);
r := numero_aleatorio(rango_r);
C := B′(ratio, T, σ, r);
return;

5.3. k-fold cross validation

En nuestro caso utilizaremos 8-fold cross validation porque nos garantizará alta varianza y bajo sesgo ..dar una explicación para estimar los hiperparámetros que definen la red neuronal que queremos construir. Los hiperparámetros propuesto para aplicar 8-fold cross validation están definidos en la Tabla 5.3. A tales fines, primero definiremos la estructura de la red neuronal entre 1 y 10 capas ocultas, variando entre 50 a 1000 neuronas por capa oculta. Luego eligiremos la función de activación [21], la inicialización de pesos de la red [22] y el algoritmo de optimización del decenso del gradiente [23] en simultaneo (haciendo todas las convinaciones posibles). Siguiendo por determinar la mejor función de error [24]. Luego determinaremos el dropout. Y por ultimo el tamaño del batch.
Para iniciar la búsqueda de los hiperparámetros óptimos, se usarán los hiperparámetros por defecto que utiliza Keras Sequential [25], excepto el tamaño del batch que será de 1024, tal como se observa en la tabla 5.2.

Cuadro 5.2: Hiperparametros por defecto

Parametros	Opciones
Función de error	ECM
Función de activación	ReLu
Inicialización de pesos	glorot_uniform
Algoritmo de optimización	SGD
Dropout	0
Tamaño de batch	1024
Learning rate	0.001
epocas	200

..motivar la introducción de estas definiciones y su uso
Las siguientes ecuaciones corresponden a métricas que miden el error de modelos. En nuestro caso se utilizará para medir la efectividad o el error de los modelos propuestos, ya sea modelos matemáticos o redes neuronales.

Cuadro 5.3: Estimación de hiperparametros

Parametros	Opciones o Rango
Capas	[1,10]
Neuronas	[50,1000]
Función de error	ECM, EAM, EPAM
Función de activación	ReLu, Elu, tanh
Inicialización de pesos	uniform, glorot_uniform, he_uniform
Algoritmo de optimización	SGD, RMSprop, Adam
Dropout	[0,0.2]
Tamaño de batch	[256, 2048]

$$\text{ECM} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{5.2}$$

$$\text{EAM} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{5.3}$$

$$\text{EPAM} = 100\frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i} \tag{5.4}$$

$$\overline{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{5.5}$$

$$SS_{tot} = \sum_{i=1}^{N}(y_i - \overline{y})^2 \tag{5.6}$$

$$R^2 = 1 - \frac{\text{ECM}}{SS_{tot}} \tag{5.7}$$

donde y_i es el valor esperado de salida de la red, $\hat{y}_i$ es el valor que predice la red y N es el tamaño de muestra.
En Cuadro 5.4 se observan los hiperparámetros obtenidos al aplicar 8-fold cross validation, ..resultantes de la optimización??? cross validation?? que se utilizará en el entrenamiento de la red.

Learning Rate

En la busqueda de los hiperparámetros ..cuándo? no se entiende hemos utilizado un **learning rate** fijo(10^{-3}).
Tomando los hiperparámetros del Cuadro 5.4 utilizaremos un método similiar

Cuadro 5.4: Hiperparametros óptimos..algo

Parametros	Opciones
Capas	3
Neuronas	950
Función de error	ECM
Función de activación	ReLu
Inicialización de pesos	random_uniform
Algoritmo de optimización	Adam
Dropout	0
Tamaño de batch	1024

al de Smith [15] para determinar el learning rate, a diferencia del método de Smith, iremos subiendo el learning rate exponencialmente por cada batch. Empezaremos con un learning rate de 10^{-10} hasta llegar a 1, comparando ECM contra el learning rate.
Como se puede observar en la Figura 5.1, el rango de learning rate óptimo se encuentra entre 5×10^{-6} y 5×10^{-3}.

Ahora proponemos tres métodos de decrecimiento del learning rate. Utilizaremos grid-search [26] ..falta referencia para encontrar los parámetros óptimos de los algoritmos de decrecimiento del learning rate. Tomando los hiperparámetros del Cuadro 5.4. En el Cuadro 5.5 se definen los intervalos sobre los cuales aplicaremos grid-search, usando como referencia el método de Smith. Dando como resultado el Cuadro 5.6.

Cuadro 5.5: Grid-Search

Parametros	Step Decay	Exponential Decay	Time-Based Decay
base_lr	$[10^{-2}, 10^{-4}]$	$[10^{-2}, 10^{-4}]$	$[10^{-2}, 10^{-4}]$
decay	[0.9, 0.95]	[0.007, 0.002]	[0.01, 8]
epoch_drop	5, 10, 20, 40, 50	-	-

Observando los parámetros del Cuadro 5.6 entrenaremos la red por 1000 épocas ulizando los algoritmos de decrecimiento del learning rate y los hiperpárametros del Cuadro 5.4 antes mencionados. Los vamos a comparar mediante su ECM, para obtener el mejor algoritmo de decrecimiento. Como se puede observar en la Figura 5.2 Step Decay es el algoritmo que mejor resultado obtuvo.

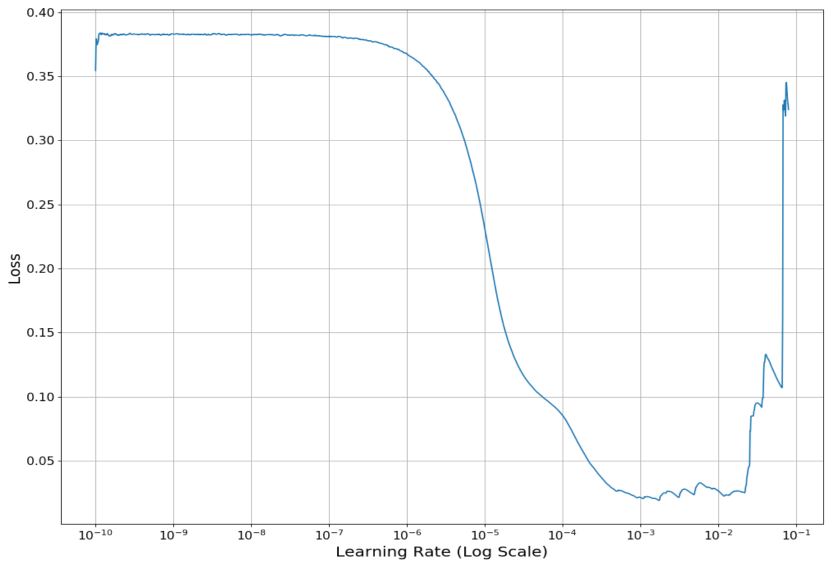

Figura 5.1: Método de Smith

Cuadro 5.6: Parametros de los Algoritmos de Decrecimiento

Parametros	Step Decay	Exponential Decay	Time-Based Decay
base_lr	$5 * 10^{-4}$	0.0005	0.005
decay	0.9	0.002	0.875
epoch_drop	20	-	-

Luego comparamos Step Decay con Cyclical Decay [15], donde el learning rate varia entre un máximo de 5×10^{-3} y un mínimo de 5×10^{-6}, con un paso cada 8 epocas aproximadamente (ver Figura 2.11). Observando la Figura 5.3, podemos concluir que Step Decay da un mejor resultado.

5.4. Optimización

Notar que si σ es grande, la derivada de la call respecto de σ es muy pequeña, no se entiende, explicar o mostrar figura (PREGUNTAR) esto puedo ocasionar un problema de gradiente pronunciado. Por lo tanto se propone un aplanamiento del gradiente para manejar este problema.
Primero, cada opción puede ser dividida entre el valor intrínseco y un valor

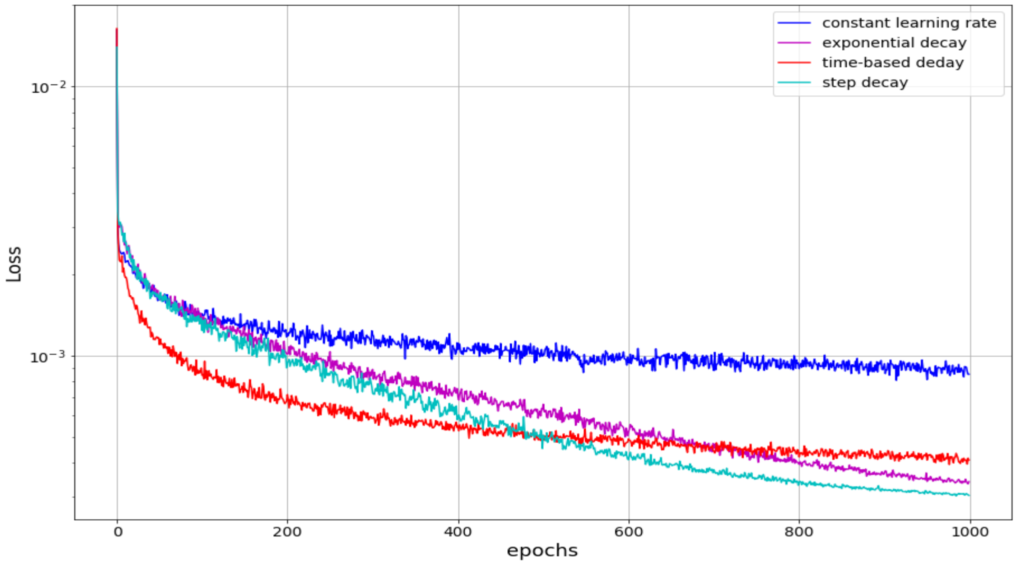

Figura 5.2: Learning Rate Decay

de tiempo, luego sustraemos el valor intrínseco de la siguiente manera:
..por qué se cambia a C?, antes se usaba c
$\tilde{c} = c - max(S_0 - Ke^{-r\tau}, 0)$
El nuevo cálculo propone superar el problema, logrando reducir el gradiente pronunciado aplicando una transformación logarítmica sobre el valor de la opción [27]. En nuestro caso sería:

$$\frac{\tilde{c}}{K} = \frac{c}{K} - \text{máx}(S/K - e^{-r\tau}, 0) \tag{5.8}$$

..ratio??

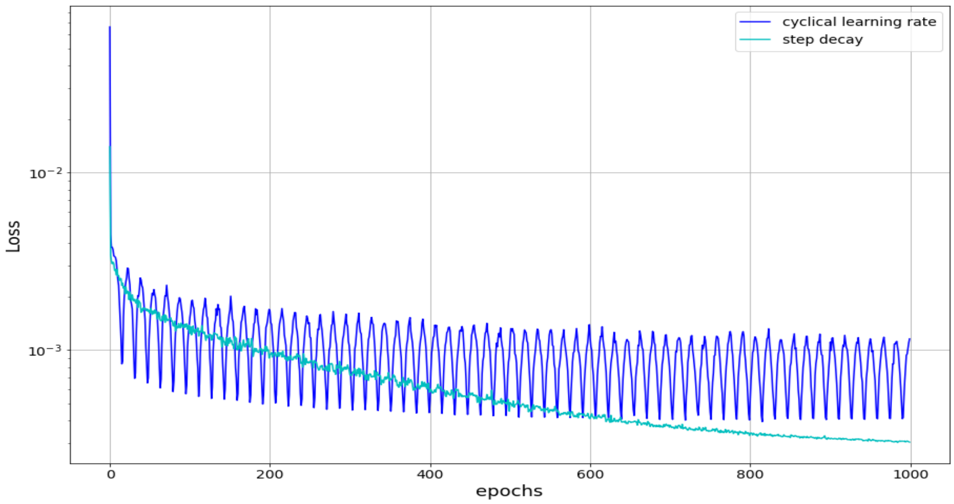

Figura 5.3: Cyclical LR vs Step Decay

Capítulo 6

Resultados

En este capítulo vamos a analizar la efectividad de los modelos propuestos. El experimiento se realizó en CPU(Intel(R) Core(TM) i5-8265U CPU @ 1.60GHz)

6.1. Primera Red

Ya tenemos los hiperparámetros óptimos del capítulo anterior, entonces vamos a definir la muestra de entrenamiento, validación y test. Crearemos 2 muestras, una amplia y una estrecha. En el cuadro 6.1 se define el ambiente de las muestras, y para generar la muestra se usa el algoritmo 5.1.

Cuadro 6.1: Hiperparametros de la Muestra

	Parametros	muestra amplia	muestra estrecha
Entrada	precio ratio(S_0/K)	[0.4, 1.6]	[0.5, 1.5]
	Tiempo de madurez(τ)	[0.2, 1.1]	[0.3, 0.95]
	volatilidad(σ)	[0.01, 1]	[0.02, 0.9]
	Tasa libre de riesgo(r)	[0.02, 0.1]	[0.03, 0.08]
Salida	Precio de Call(c/K)	(0, 0.9)	(0, 0.73)

Una vez obtenida la muestra, la entrada de la red será {c/K, S_0/K, r, τ} y la salida {σ}. Vamos a entrenar la red durante 1000 épocas utilizando la muestra amplia y los hiperparámetros del obtenidos capítulo anterior.
Una vez entrenada la red, en el Cuadro 6.2 y en las Figuras 6.1 y 6.2 se pueden observar los resultados.

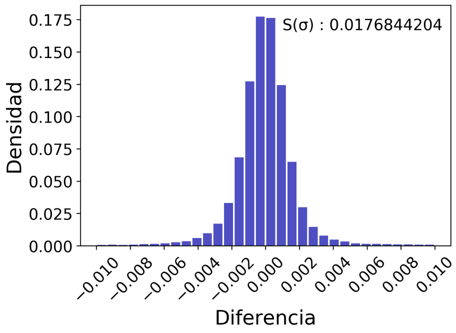

Figura 6.1: Predicción muestra amplia.

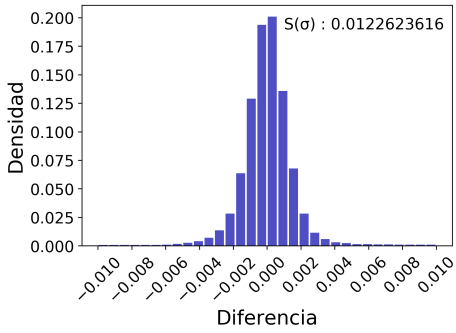

Figura 6.2: Predicción muestra estrecha.

Cuadro 6.2: Error Red

	ECM	EAM	EPAM	R^2
Muestra Amplia	$3{,}21e(-4)$	$6{,}09e(-3)$	9,30	0.996031
Muestra Estrecha	$1{,}47e(-4)$	$4{,}07e(-3)$	5,46	0.997714

6.2. Optimización

Ya tenemos los hiperparámetros óptimos del capítulo anterior, entonces vamos a definir la muestra de entrenamiento, validación y test. Crearemos 2 muestras, una amplia y una estrecha. A diferencia de la muestra de la sección anterior, en esta se va a modificar utilizando la ecuación 5.1 del cápitulo anterior. En el cuadro 6.3 se define el ambiente de la muestra, y para generar las muestras se usa el algoritmo 5.1.

Cuadro 6.3: Hiperparametros de la Muestra

	Parametros	muestra amplia	muestra estrecha
Entrada	precio ratio(S_0/K)	[0.4, 1.6]	[0.5, 1.5]
	Tiempo de madurez(τ)	[0.2, 1.1]	[0.3, 0.95]
	volatilidad(σ)	[0.01, 1]	[0.02, 0.9]
	Tasa libre de riesgo(r)	[0.02, 0.1]	[0.03, 0.08]
Salida	Precio de Call($\log(\tilde{C}/K)$)	[-16.12,-0.94]	[-16.12,-0.94]

Una vez obtenida la muestra, la entrada de la red será $\{\log(\tilde{C}/K)$, S_0/K, r, $\tau\}$ y la salida $\{\sigma\}$. Vamos a entrenar la red durante 1000 épocas.

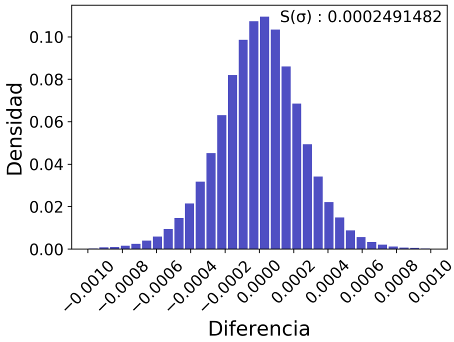

Figura 6.3: Predicción muestra amplia.

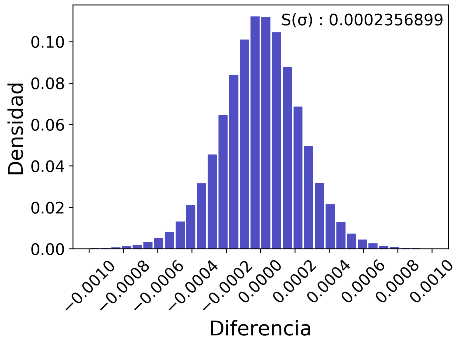

Figura 6.4: Predicción muestra estrecha.

Una vez entrenada la red, en el Cuadro 6.4 y en las Figuras 6.3 y 6.4 se pueden observar los resultados.

Cuadro 6.4: Error Red Optimizada

	ECM	EAM	EPAM	R^2
Muestra Amplia	$6{,}24e(-8)$	$1{,}92e(-4)$	0,059	0.999999104
Muestra Estrecha	$5{,}58e(-8)$	$1{,}84e(-4)$	0,062	0.999999020

6.2.1. Cambio en función de decrecimiento del Learning Rate

Observando la Figura 5.1 podemos concluir que el entrenamiento de la red con un learning rate menor a 10^{-7} prácticamente no actualizaría los pesos, entonces si queremos hacer mas épocas en el entrenamiento de la red, con el algoritmo que usamos en la sección anterior tendríamos en la época 3000 un learning rate aproximado de $1{,}87e^{-17}$, y en la época 1000 el learning rate aproximado sería de $2{,}65e^{-8}$. Una posible solución a este problema es modificar los hiperparámetros óptimos aumentando el *base_lr*, aumentar el *decay*, y aumentar el *epoch_drop*, para que el learnig rate descienda mas lentamente. Otra solución sería cuando el learning rate sea menor a un umbral, aumentar el *base_lr*. El umbral puede ser variable.
El algoritmo 6.1 muestra el método de decrecimiento de learning rate que utilizaremos:

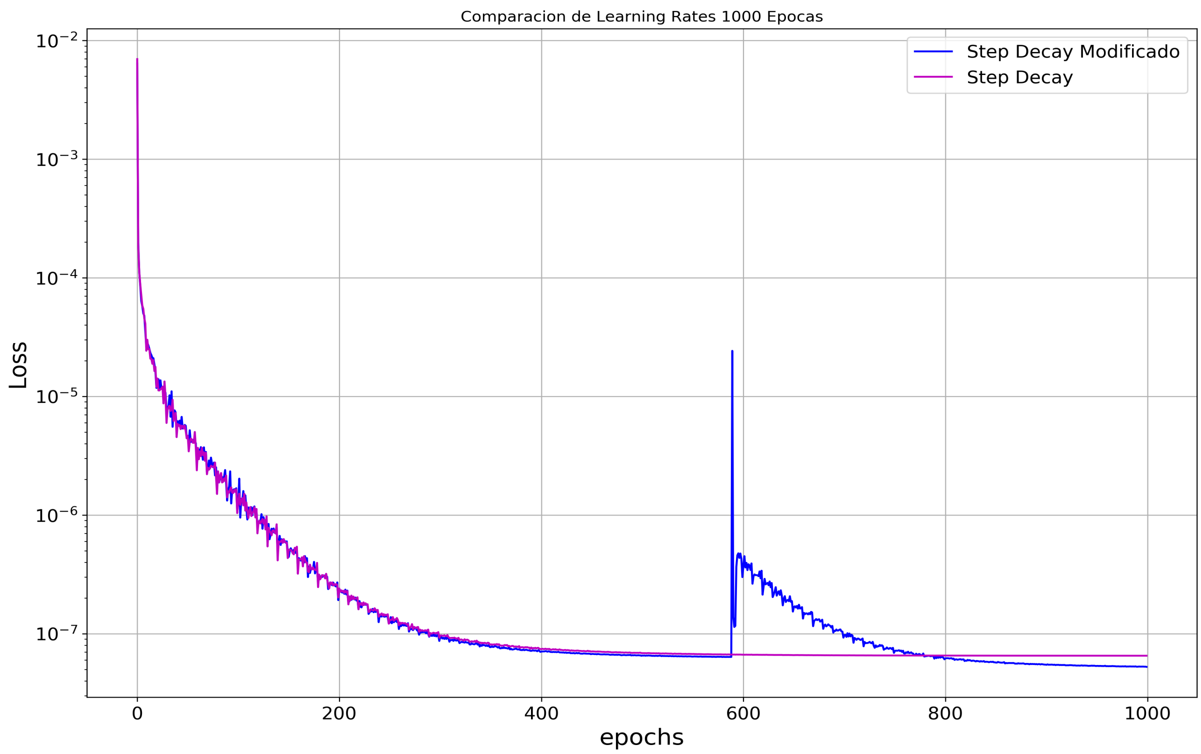

Figura 6.5: Step Decay Modificado

Algorithm 6.1: Step Decay Modificado

Input : epoch
Output: lrate

i := 0;
step := floor((1+epoch)/10);
lrate := 0.0005 * pow(0.9, step);
while $lrate < 10^{-(6+i)}$ **do**
    lrate := 100*lrate;
    i = i + 0.5;
end
return;

La Figura 6.5 muestra la comparación entre el step decay usado anteriormente y el step decay modificado.
Las Figuras 6.6 y 6.7 muestan el learning rate resultante con los diferentes algoritmos.

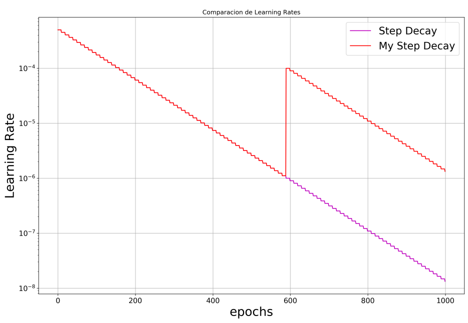

Figura 6.6: Learning Rate sobre 1000 epocas.

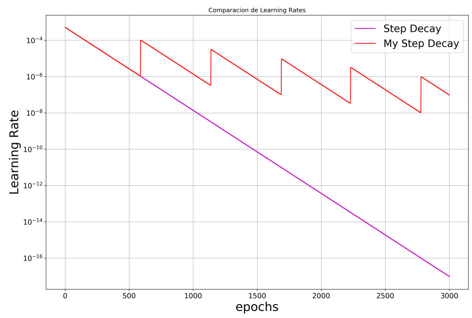

Figura 6.7: Learning Rate sobre 3000 epocas.

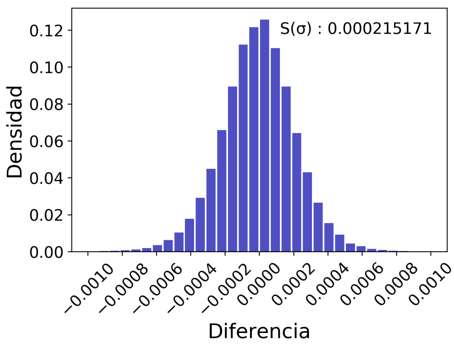

Figura 6.8: Predicción muestra amplia.

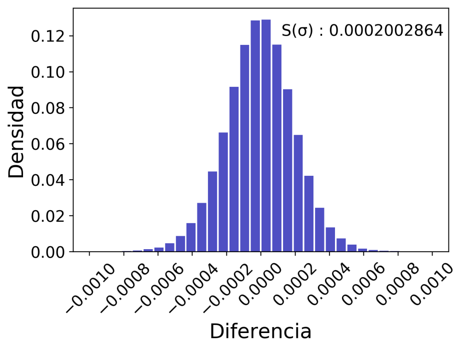

Figura 6.9: Predicción muestra estrecha.

Luego vamos a entrenar la red durante 3000 epocas usando step decay modificado. En las Figuras 6.8, 6.9, y en el Cuadro 6.5 se muestran los resultados.

Cuadro 6.5: Error Red Optimizada

	ECM	EAM	EPAM	R^2
Muestra Amplia	$4{,}67e(-8)$	$1{,}66e(-4)$	0,0515	0.999999329
Muestra Estrecha	$4{,}02e(-8)$	$1{,}57e(-4)$	0,0526	0.999999295

6.3. Métodos Numéricos

6.3.1. Problemas Númericos

En la muestra amplia definida en la sección 6.1 ..muestra (NO se entiende porque no fue definida la muestra. Tal vez conviene llevar este capítulo después del siguiente o adelantar la definición de muestra) hay aproximado un 0.05 % de casos que no cumplen con la condición $g(a)g(b) < 0$, para $a \to 0$ y $b \to \infty$, donde g es la función 4.1. ..Usar siempre label y ref Por lo tanto, no se puede aplicar ninguno de los métodos vistos anteriormente.
El problema radica en el cálculo de la fórmula de Black-Scholes definida en la ecuación 2.1 ..NO –¿el capítulo 2.1 (ecuación 2.1). Usar las referencias de latex, básicamente en la distribución de probabilidad acumulada $\Phi(d)$. Por definición $\Phi(d) < 1$, $\forall d \in \mathbb{R}$. Pero Python usa aritmética de punto flotante IEEE-754 [19], donde su precisión es $2^{(-56)}$ ..falta referencia , dando asi una precisión aproximada de 15,95 dígitos decimales. Luego usando la función de Scipy para el cálculo de la distribución normal acumulada definida en la referncia [18], obtenemos $\Phi(n) = 1$, para $n > 8{,}3$. Volviendo a la fórmula de Black-Scholes, si bien sabemos que $\Phi(d_1) > \Phi(d_2)$, ya que ..o ??, $\sigma, T > 0$, pero hay casos en que el extremo inferior definido para aplicar el método de bisección (o Brent), se obtiene $\Phi(d_1) = \Phi(d_2)$ por el problema de precisión, ocasionando el error antes mencionado.
El origen de todo el problema radica en la cantidad de posibles números representables en un dado rango. El formato se escribe con un significando que tiene un bit entero implícito de valor 1 (excepto para los números especiales). Con los 52 bits de la mantisa, la precisión total es por lo tanto de 53 bits (es decir de $53 \log_{10}(2) \approx 15{,}955$ que se redondea a 16 dígitos decimales). El exponente de este formato está sesgado o desplazado en 1023 unidades, ya que como el máximo valor representado por 11 bits es $2^{11} - 1 = 2047$, es la mitad de este rango la que representa exponentes positivos y la otra, exponentes negativos [19]. Podemos observar esto en la Figura 6.10. Lo expuesto permite ver que hay 2^{52} números de punto flotante entre 2^n y 2^{n+1} para todo n entero en $[-1023, 1022]$.
Por ejemplo, consideremos un caso en particular, veamos los valores del Cuadro 6.6
..llevar esto a una tabla con label y caption
La Figura 6.11 muestra la comparación entre $g(n)$ para $n \in [0{,}01, 1]$ y la volatilidad implícita (linea roja punteada). Se puede observar que no se puede hallar el intervalo $[a, b]$ para aplicar el método de bisección (o Brent), ya que para este caso $g(\sigma)$ es positivo $\forall \sigma$, tal que $0 < \sigma < 0{,}01$.

Figura 6.10: Estructura de un número en formato de coma flotante de doble precisión

c:	5.983489610184446
S:	15.752756180327959
k:	10
r:	0.09010364215460305
T:	0.2590760904347537

Cuadro 6.6: Parámetros que no cumplen la pre-condición de los métodos numéricos

Tomando $\sigma = 0{,}01$ (extremo inferior del intervalo), se obtiene, $\Phi(d_1) = 1$, $\Phi(d_2) = 1$ y $g(0{,}01) = 1{,}7763568394002505e(-15)$. Pero si se evalúa la función de distribución acumulada de la normal estándar utilizando la volatilidad implícita $\sigma = 0{,}11928197090875538$, se obtiene $\Phi(d_1) = 0{,}9999999999999986$, $\Phi(d_2) = 0{,}9999999999999982$ y $g(0{,}11928197091) = 0$, se contradice la propiedad de monotonía creciente.
..Falta describir en detalle las dos últimas figuras en el texto y citarlas En esta sección también falta la discusión que tuvimos con las gráficas de la prima en función de la volatilidad y el efecto de la vega. La discusión que está en los correos de comienzos de junio.

6.3.2. Prescisión Métodos Numéricos

En la sección 6.3.1 vimos que hay casos en los que no se puede aplicar los métodos numéricos, exceptuando esos casos, vamos a analizar su precisión. Ademas se vió que lo máximo que podemos aspirar es un error absoluto de 2^{-56}. Ahora vamos a analizar si en la práctica, la tolerancia de los métodos corresponden con los resultados obtenidos de aplicar los mismos. En el método de Brent utilizamos el de la librería Scipy. [28]
Observando el Cuadro 6.7, el error absoluto medio del método de bisección es $2{,}89e(-6)$ y de brent es $9{,}18e(-5)$, siendo la tolerancia $2e(-56)$, esto es por el problema de presición y el impacto que tiene la volatilidad sobre el precio de la opción (vega) utilizando la fórmuala de Black Scholes 2.1. En la Figura 6.14, que compara el precio de la opción con respecto a la variación de

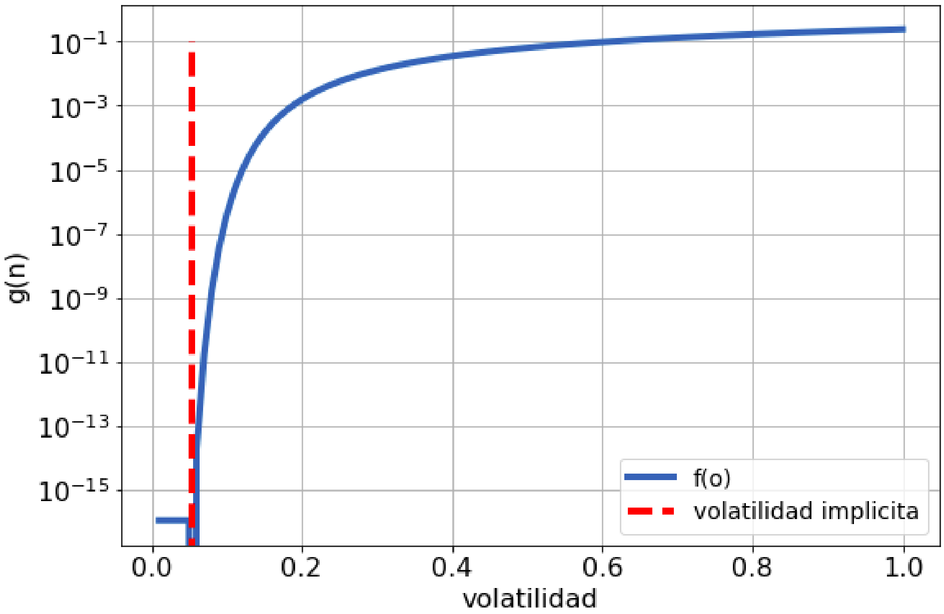

Figura 6.11: Función g ..aumentar el tamaño de font y ancho de la línea)

Cuadro 6.7: Error brent y Bisección

	ECM	EAM	EPAM	R^2
Bisección	1,10$e(-8)$	2,89$e(-6)$	0,00469	0.99999986
Brent	8,17$e(-6)$	9,18$e(-5)$	0,35894	0.99989551

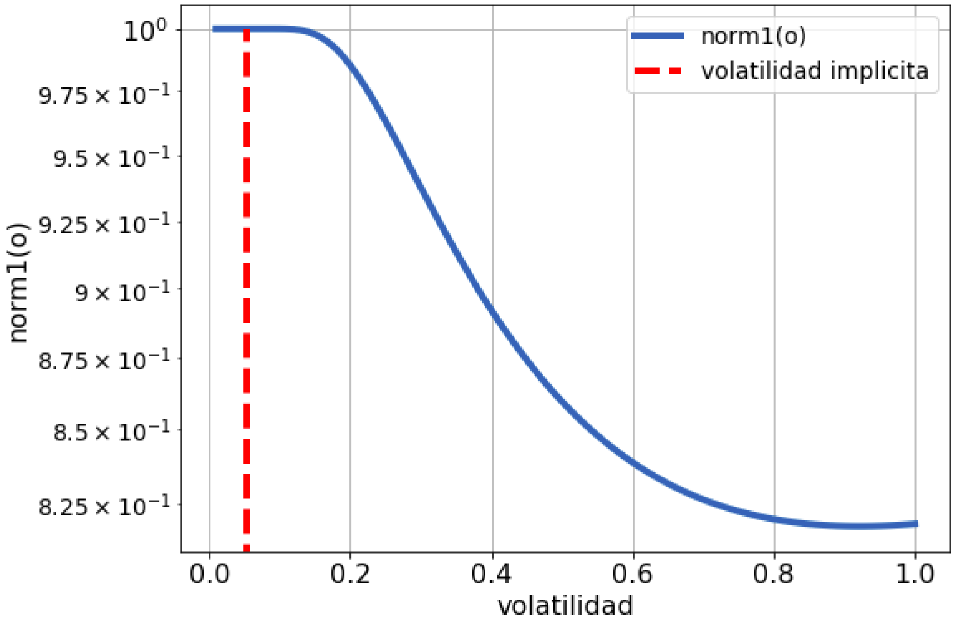

Figura 6.12: $\Phi(d_1)$. ..aumentar el tamaño de font y ancho de la línea)

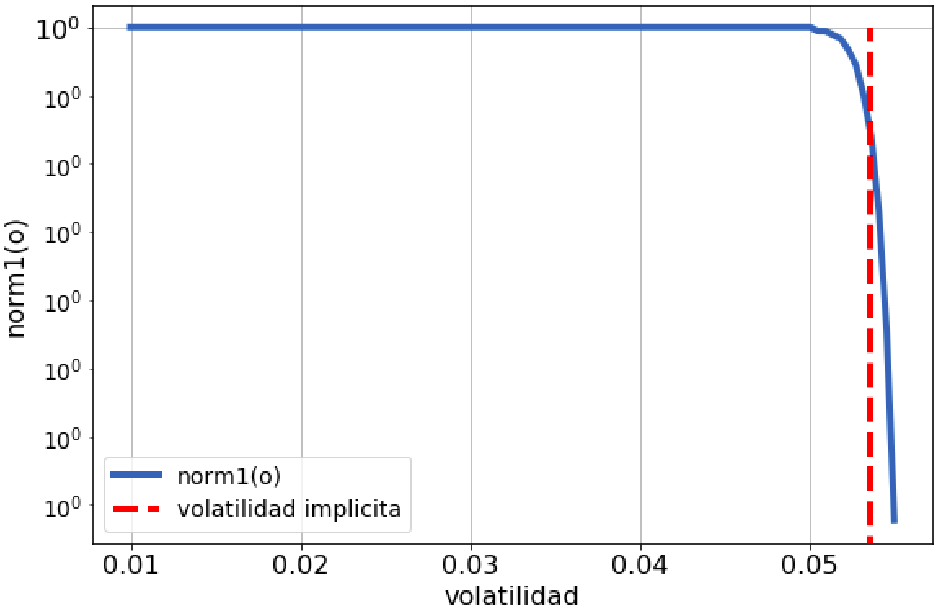

Figura 6.13: $\Phi(d_1)$. ..aumentar el tamaño de font y ancho de la línea)

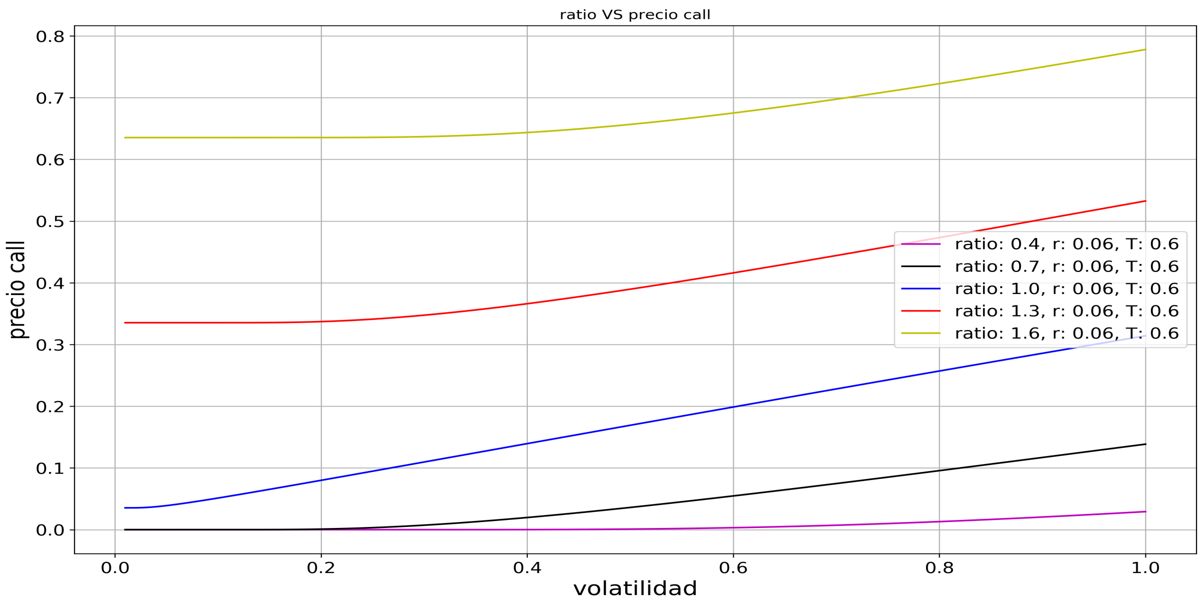

Figura 6.14: Volatilidad implícita vs precio call con distiontos ratios

volatilidad, utilizando diferentes ratios (S/K), se puede observar que mientras más diferencia hay entre S_0 y K la volatilidad tiene menos impacto sobre el precio en especial en los casos OTM y en las volatilidades cercanas a 0. Luego en la Figura 6.15, que compara el precio de la opción con respecto a la variación de volatilidad, utilizando diferentes tasas de interés, se observa practicamente un desplazamiento. Por último en la Figura 6.16, que compara el precio de la opción con respecto a la variación de volatilidad, utilizando diferentes tiempos de madurez, se puede observa que mientras los contratos sean más largos la volatilidad tiene un mayor impacto sobre el precio.
Ahora bien observando los gráficos podemos concluir que la volatilidad tiene menos impacto en los contratos cortos(en nuestro ambiente), en OTM y con volatilidades menores a 0.2, pero en la practica los casos con mas error son ITM con $\Phi_1(\sigma)$ y $\Phi_2(\sigma)$ cercanos a 1.
Esto se debe a la densidad de puntos flotates visto en el capítulo 4. Luego como el ancho de 11 bits del exponente permite la representación de números en el rango comprendido entre 2^{-1023} y 2^{1023} (10^{-308} y 10^{308}). Esto nos permite concluir que hay la misma cantidad de numeros representables entre 10^{-308} y 10^{-307} que en 0,1 y 1, osea que su densidad es mayor cuando tiende a 0 (Observar Figura 6.17). Por lo tanto los $\Phi_1(\sigma)$ y $\Phi_2(\sigma)$ cercanos a 0 en OTM van a tener mas precisión aunque el peso de la volatilidad sobre el precio de la opción sea menor.

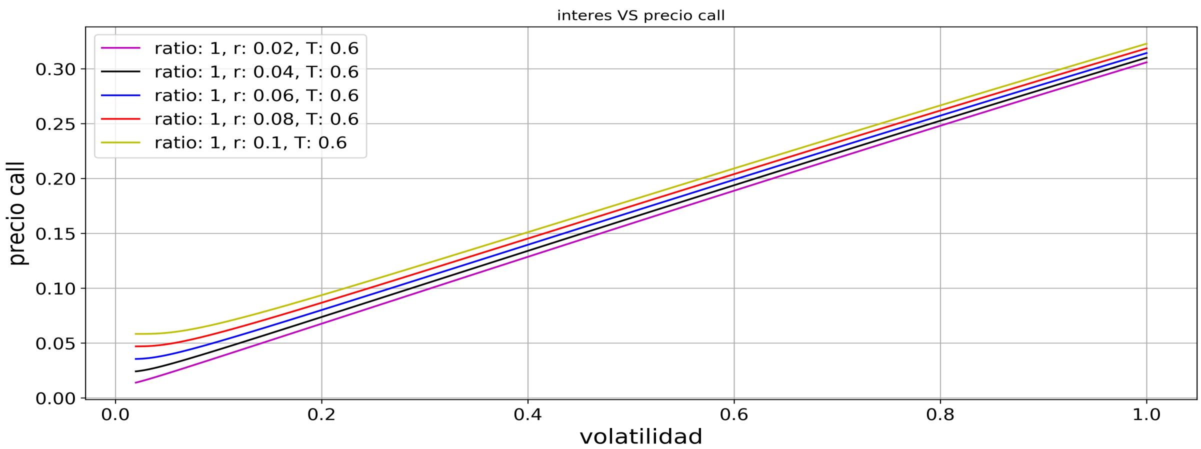

Figura 6.15: Volatilidad implícita vs precio call con distiontas tasas de interés

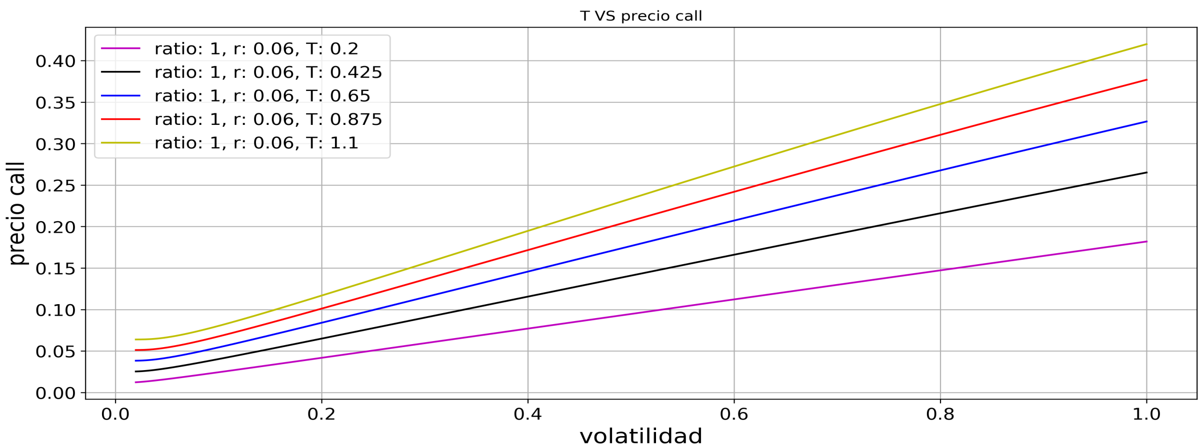

Figura 6.16: Volatilidad implícita vs precio call con distiontos tiempos de madurez

6.4. Tiempo de Ejecución y Robustez

Una vez entrenada la red (Fue calculado sobre GPU(GeForce GTX 1080/PCIe/SSE2), RAM(15.6 GiB). La busqueda de hiperpámetros en un tiempo 10.08 horas, la busqueda del algoritmo de learning rate fue 11.72 horas y entrenamiento de la red 3000 epocas fue 8.83 horas), vamos a observar el tiempo ejecución de evaluación de la misma sobre una muestra de 10000. Ademas el tiempo de ejecución de los métodos de brent y bisección sobre la misma muestra en

Figura 6.17: Densidad de números representables en rango de 10^{-5} entre 0.1 y 1

una CPU(Intel(R) Core(TM) i5-8265U CPU @ 1.60GHz).
Tambien analizaremos su robustez, esto sería, si para todo elemento de la muesta el modelo genera una salida.
En el Cuadro 6.8 se observan el tiempo de ejecución y la Robustez.

Cuadro 6.8: Robustez y Tiempo de Ejecución

	Red Neuronal	Método de Brent	Método de Bisección
Tiempo de Ejecución(segundos)	7,34	157,05	318,31
Robustez	Si	No	No

Capítulo 7

Conclusiones

En este trabajo se presentó la importancia del cálculo de la volatilidad implícita. Ademas se propuso una manera diferente del cálculo que usualmente se utiliza, usando redes neuronales feedforward.
Si bien el método de bisección estima mejor la volatilidad implícita, no es una mejora significativa con respecto a la red neuronal o al método de brent, sin embargo el tiempo de ejecución la red neuronal es 21.12 veces mas rápido que el método de Brent y 43.37 veces mas rápido que el método de bisección.
En caso de tener datos atípicos, por ejemplo precio de strike tres veces mas grande que el precio del subyancente, o tasas libre de riesgo superiores al 20 %, o contratos muy largos, en esos casos es conveniente usar los métodos numéricos, ya que la red no va a estimar bien, porque la red no entrenó con esos datos. Sin embargo hay casos en los que los métodos numéricos no pueden ser aplicados.
En caso que la presición sea imprescindible, entonces se utilizará el método de bisección.

Dentro de las posibilidades de continuar con el trabajo tenemos, uso de GPU en la implementación de los modelos matématicos propuestos, presentar soluciones para el problema numérico del capítulo 4.
En este trabajo se planteó el cálculo de la volatilidad implícita mediante opciones call europeas, ademas se podrían proponer otros modelos obtener la volatilidad impícita mediante puts europeas.
La bibliografía hay que normalizarla con alguno de los estándar en uso. El más próximo a lo que tenés es numeric (ver: https://www.overleaf.com/learn/latex/biblatex_bibliograp
y completar todos los campos. El los artículos simpre usar nombre de journal, vol, pag (año). En los libros siempre debe ponerse la editorial. Cuidar que el orden sea el de cita en el texto.

Bibliografía

[1] Patricia Kisbye. Modelos matemáticos en finanzas cuantitativas. 2019.

[2] Mark Yallop. The future of the otc markets. `https://thehill.com/sites/default/files/ICAP_TheFutureoftheOTCMarkets_0.pdf`, 2008.

[3] Jay L. Devore. *Probabilidad y Estadística para Ingeniería y Ciencias*, páginas 144–148. Cengage Learning, séptima edición, 2009.

[4] John C. Hull. *Options, Futures, and Other Derivatives*, páginas 381–387. Pearson Prentice Hall, séptima edición, 2009.

[5] John C. Hull. *Options, Futures, and Other Derivatives*, páginas 387–389. Pearson Prentice Hall, séptima edición, 2009.

[6] Ward Cheney y David Kincaid. *Métodos numéricos y computación*, páginas 78. Cengage Learning, sexta edición, 2011.

[7] Ward Cheney y David Kincaid. *Métodos numéricos y computación*, páginas 76–82. Cengage Learning, sexta edición, 2011.

[8] Ward Cheney y David Kincaid. *Métodos numéricos y computación*, páginas 111–112. Cengage Learning, sexta edición, 2011.

[9] William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery. *Numerical Recipes The Art of Scientific Computing*. páginas 454–456. Cambrindge University Press, tercera edición, 2007

[10] Tom M. Mitchell. *Machine Learning*. páginas 86–88. McGraw-Hill Education, 1997

[11] John A. Hertz, Anders S. Krogh, Richard G. Palmer. *Introduction To The Theory Of Neural Computation*. páginas 90–92. Westview Press, 1991.

[12] Wei Di, Anurag Bhardwaj, Jianing Wei. *Deep Learning Essentials*. páginas 61–62. Packt Publishing, 2018.

[13] Jason Brownlee. *Deep Learning With Python*. páginas 109–112. (2016),

[14] Wei Di, Anurag Bhardwaj, Jianing Wei. *Deep Learning Essentials*. páginas 219 Packt Publishing, (2018).

[15] Leslie N. Smith. Cyclical Learning Rates for Training Neural Networks. *ArXiv:abs/1506.01186*. `https://arxiv.org/abs/1506.01186`, 2015.

[16] Prosper Lamothe Fernández, Miguel Pérez Somalo. *Opciones Financieras y Productos Estructurados*. páginas 319–336. McGraw-Hill Interamericana de España S.L., 2003.

[17] Jay Kaeppel. *The Option Trader's Guide to Probability, Volatility, And Timing*. páginas 156–201. John Wiley & Sons, Inc, 2002

[18] Distribución Normal Acumulada. `https://docs.scipy.org/doc/scipy-0.16.0/reference/generated/scipy.stats.norm.html`.

[19] Densidad Punto Flotante. `https://web.archive.org/web/20160806053349/http://www.csee.umbc.edu/~tsimo1/CMSC455/IEEE-754-2008.pdf`.

[20] Ramazan Gençay René Garcia. Pricing and hedging derivative securities with neural networks and a homogeneity hint. *Journal of Econometrics*, 94:93–115, Enero 2000.

[21] Función de Activación. `https://keras.io/activations/`

[22] Inicialización de Pesos. `https://keras.io/initializers/`

[23] Optimizadores del Descenso del Gradiente. `https://keras.io/optimizers/`

[24] Función de Error. `https://keras.io/losses/`

[25] Modelo Secuencial. `https://keras.io/api/models/sequential/`

[26] Grid Search. `https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html#sklearn.model_selection.GridSearchCV`

[27] Shuaiqiang Liu, Cornelis Oosterlee, and Sander Bohte. Pricing options and computing implied volatilities using neural networks. *Risks*, 7(1):16, Feb 2019.

[28] Método de Brent. `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.brentq.html`