

## Práctica 3

# Descenso del Gradiente y Combinador Lineal



### Objetivos

El objetivo de esta práctica es comprender el funcionamiento del descenso del gradiente y los combinadores lineales.

### Temas

- Optimización con descenso del gradiente.
- Regresión lineal simple y múltiple.
- Regresión logística multiclase

### Lectura

Material de Lectura: Capítulos 12 y 13 del libro Master Machine Learning Algorithms.

## Optimización de funciones utilizando la técnica del descenso de gradiente

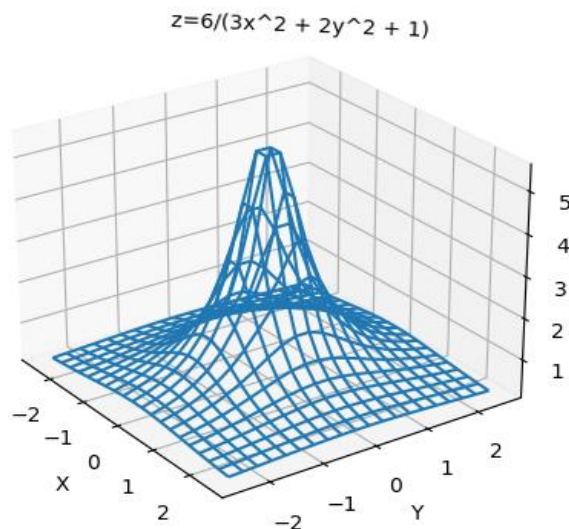
### Ejercicio 1

Dada la función  $f(x, y) = 6/(3x^2 + 2y^2 + 1)$  cuyas derivadas parciales son:

$$\frac{\partial f(x, y)}{\partial x} = -36x/(3x^2 + 2y^2 + 1)^2$$

$$\frac{\partial f(x, y)}{\partial y} = -24y/(3x^2 + 2y^2 + 1)^2$$

- Indique el vector gradiente para  $x = 1$  e  $y = -1$
- Si a partir de  $(x = 1; y = -1)$  tuviera que elegir una dirección en la cual la función crezca, qué acción tomaría?
  - Incrementar los valores de  $x$  e  $y$
  - Disminuir los valores de  $x$  e  $y$
  - Incrementar el valor de  $x$  y disminuir el de  $y$
  - Disminuir el valor de  $x$  e incrementar el de  $y$
- Una vez elegida la dirección en b), el movimiento a realizar será una fracción del módulo del vector gradiente. ¿Cuál sería la nueva posición si se considera  $\alpha=0.1$ ? ¿cuál es el valor de la función en la nueva ubicación?
- Considere nuevamente las opciones del inciso b) pero ahora debe buscar una dirección, a partir de  $(x = 1; y = -1)$ , en la cual la función disminuya ¿cuál sería su elección?
- Repita lo solicitado en c) para la dirección elegida en d) donde el objetivo fue hallar una ubicación donde la función tome un valor menor.



### Ejercicio 2

Utilice la técnica del descenso de gradiente para calcular el mínimo de las siguientes funciones:

$$a) V(r, h) = r^2 h \pi / 3 \quad ; \quad \frac{\partial f(r, h)}{\partial r} = 2rh\pi/3 \quad ; \quad \frac{\partial f(r, h)}{\partial h} = r^2\pi/3$$

$$b) f(x, y) = -3/(x^2 + y^2 + 1) \quad ; \quad \frac{\partial f(x, y)}{\partial x} = 6x/(x^2 + y^2 + 1)^2 \quad ; \quad \frac{\partial f(x, y)}{\partial y} = 6y/(x^2 + y^2 + 1)^2$$

### Ejercicio 3

El desempeño de la técnica de optimización basada en la dirección del gradiente depende, en gran medida, del valor del factor *alfa* que indica cuánto del módulo del vector gradiente se utilizará para modificar los parámetros actuales de la función con el objetivo de avanzar en la dirección del óptimo.

- En referencia a la función  $V(r, h)$  del ejercicio 2.a), utilizando en todos los casos  $\text{MAX\_ITE} = 100$  y una cota de  $10e - 06$  para el valor absoluto de la diferencia de dos valores consecutivos de la función, indique cuáles de las siguientes configuraciones logra encontrar una ubicación donde  $V(r, h)$  vale aproximadamente 0 (cero). En caso de hallar la ubicación donde la función vale cero, informe los valores de  $r$  y de  $h$  y en caso de no hallarla, analice los motivos.

- i. Punto de inicio:  $(r = 0.5, h = -0.5)$  ;  $\alpha = 0.01$
  - ii. Punto de inicio:  $(r = 0.5, h = 0.5)$  ;  $\alpha = 0.5$
  - iii. Punto de inicio:  $(r = 1.5, h = 1.5)$  ;  $\alpha = 0.5$
- b) En referencia a la función  $f(x, y)$  del ejercicio 2.b), utilizando en todos los casos  $\text{MAX\_ITE} = 100$  y una cota de  $10e - 06$ , indique cuáles de las siguientes configuraciones logra hallar la ubicación del mínimo:
- i. Punto de inicio:  $(x = 0.5, y = -0.5)$  ;  $\alpha = 0.001$
  - ii. Punto de inicio:  $(x = -2.0, y = 2.0)$  ;  $\alpha = 0.05$
  - iii. Punto de inicio:  $(x = 1.5, y = 1.5)$  ;  $\alpha = 0.5$

#### Ejercicio 4

El precio de una vivienda está dado por dos factores, la proximidad a centros comerciales, financieros, vías y demás que se mide a través del factor  $j$ , y de la condición misma de la vivienda que se mide por el factor  $k$ . La relación entre estos factores y el precio de la vivienda en un barrio determinado está dada por:

$$P(j, k) = k^2 - 2jk + 2j^2 + 18j - 20k + 201$$

donde el precio está dado en miles de dólares.

Utilice la técnica de descenso de gradiente para determinar los valores de  $j$  y  $k$  que minimizan el precio de la vivienda.

### Regresión Lineal Simple (una única variable de entrada o variable independiente)

#### Ejercicio 5

A partir de las siguientes calificaciones de alumnos de curso, se desea obtener la recta de regresión que permita predecir la nota que un estudiante obtendrá en Física a partir de su calificación obtenida en Matemáticas.

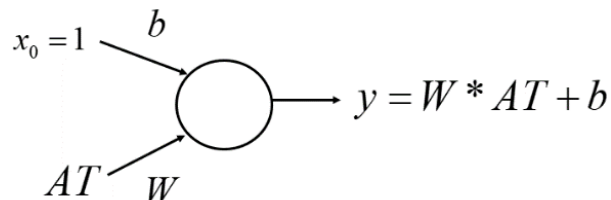
- a) Calcule la correlación lineal entre estas dos variables y dibuje el diagrama de dispersión correspondiente. ¿La recta de regresión será de utilidad?
- b) Indique la ecuación del error cuadrático medio que se debería minimizar si se utilizara la técnica de descenso de gradiente.
- c) Indique la ecuación de error que se minimiza al utilizar la técnica de descenso de gradiente estocástico. ¿qué ventaja tiene su uso con respecto a la ecuación indicada en a)?
- d) Utilice la técnica de descenso de gradiente estocástico para obtener los coeficientes de la recta de regresión que permita predecir la nota que un estudiante obtendrá en Física a partir de su calificación obtenida en Matemáticas.
- e) Utilice la recta obtenida en c) para predecir la calificación que obtendrá en Física un estudiante que obtuvo un 9 en el examen de matemáticas.

Matemáticas	Física
2	1
3	3
4	2
4	4
5	4
6	4
6	6
7	4
7	6
8	7
10	9
10	10

**Ejercicio 6**

El archivo CCPP.csv contiene 9568 datos de una central de ciclo combinado recolectados entre 2006 y 2011. Usando la totalidad de los ejemplos se entrenaron 5 combinadores lineales con el objetivo de predecir el valor del atributo PE correspondiente a la cantidad de energía producida a partir del atributo AT (más detalles de este conjunto de datos aquí). Indique cuáles son los dos combinadores lineales con menor error cuadrático promedio a la hora de predecir el valor de PE en función de AT:

- a)  $W = -2.00$  ;  $b = 500$
- b)  $W = -2.21$  ;  $b = 498$
- c)  $W = -2.30$  ;  $b = 497$
- d)  $W = -2.22$  ;  $b = 496.5$
- e)  $W = -2.16$  ;  $b = 496.91$

**Ejercicio 7**

A partir de los ejemplos del archivo AUTOS.csv, utilizado en la práctica 1, se entrenará un combinador lineal para predecir el precio del auto (atributo 'price') en función del tamaño del motor (atributo 'engine-size'):

- a) Antes de comenzar con el entrenamiento, complete los datos faltantes del atributo 'price' utilizando el valor promedio.
- b) Luego de completar según lo indicado en a), realice el entrenamiento del combinador lineal:
  - i. Sin normalizar los atributos.
  - ii. Normalizando ambos atributos linealmente.
  - iii. Normalizando ambos atributos utilizando media y desvío.

Analice el impacto de la normalización en el entrenamiento. ¿Qué resultados obtuvo sin normalizar los atributos? ¿Alguna de las normalizaciones es mejor? Explique.

**Regresión Lineal Múltiple (dos o más variables de entrada)****Ejercicio 8**

Entrene un combinador lineal que reciba tres dígitos binarios y devuelva el número decimal correspondiente.

- a. Utilizando el bias o peso  $W_0$
- b. Utilizando únicamente las tres entradas correspondientes a los dígitos binarios sin el bias.

Compare la cantidad de iteraciones necesarias para obtener el vector de pesos correcto en ambos casos. Observe el valor del arco correspondiente al bias en a).

**Ejercicio 9**

Utilizando los atributos numéricos del archivo Autos.csv entrene un combinador lineal para predecir el precio del auto en función del resto de los atributos. Tenga en cuenta la existencia de valores faltantes. Realice varias ejecuciones y observe los vectores de pesos obtenidos. ¿Puede identificar cuáles son los atributos con mayor incidencia en la determinación del precio? Explique

## Regresión Logística Múltiple (varias variables de entrada y salida sigmoide en (0,1) o (-1,1) )

### Ejercicio 10

El archivo **SEMILLAS.csv** contiene información de granos que pertenecen a tres variedades diferentes de trigo.

Entrene una neurona no lineal para determinar si una semilla es de “TIPO 2” o no.

Utilice el 60% de los ejemplos para entrenar y el 40% para realizar el testeo.

Realice 30 ejecuciones independientes de la configuración seleccionada para respaldar sus afirmaciones referidas a la precisión obtenida tanto para el conjunto de entrenamiento como para el de testeo. Utilice un máximo de 200 iteraciones y una cota de error de  $1e - 03$ .

- Utilizando ‘sigmoid’ como función de activación.
- Utilizando ‘tanh’ como función de activación.

Con los resultados obtenidos complete la siguiente tabla

Funcion	alfa	Iteraciones	Accuracy train	Ite accuracy	Veces (datos Train)	Accuracy test	Veces (Datos Test)
sigmoid	0,3						
	0,2						
	0,1						
tanh	0,3						
	0,2						
	0,1						

donde

- **Iteraciones** es la cantidad promedio de iteraciones realizadas en las 30 ejecuciones
- **Accuracy train** es el promedio de las 30 tasas de acierto obtenidas en las distintas ejecuciones.
- **Ite accuracy** es el número de iteración promedio en el que se encontró el máximo accuracy por primera vez.
- **Veces (datos train)** indica cuántas de las 30 ejecuciones alcanzaron el 100% de acierto durante el entrenamiento.
- **Accuracy test y Veces (datos Test)** contienen lo mismo que Accuracy train y “Veces (datos train)” pero sobre los datos de testeo, respectivamente.

### Ejercicio 11

Se ha realizado un análisis químico a tres tipos distintos de vinos producidos en una misma región de Italia. El número de muestras considerado es el siguiente:

- Tipo 1 → 59 muestras
- Tipo 2 → 71 muestras
- Tipo 3 → 48 muestras

El archivo **Vinos.csv** permite observar los resultados de este análisis. Cada fila representa una muestra distinta y está formada, en primer lugar, por el número del tipo al cual pertenece el vino analizado seguido por los 13 atributos que lo caracterizan.

Por ejemplo, la siguiente fila:

**2, 12.29, 3.17, 2.21, 18, 88, 2.85, 2.99, 0.45, 2.81, 2.3, 1.42, 2.83, 406**

es el resultado del análisis de un vino correspondiente al tipo 2 (1er. valor de la fila) seguido por 13 valores separados por comas que indican los niveles de las mediciones realizadas a dicho vino.

- a) Entrene una red neuronal formada por una única neurona para clasificar los vinos de Tipo 1

Realice 30 ejecuciones independientes utilizando el 50%, 60%, 70%, 80% y 90% de los ejemplos como entrenamiento y el resto como testeo. Para cada porcentaje, indique la cantidad promedio de ejemplos correctamente clasificados en entrenamiento y en testeo. Calcule también el promedio y el desvío de la cantidad de iteraciones realizadas.

Utilice un máximo de 400 iteraciones y velocidades de aprendizaje 0.1, 0.2 y 0.3.

Analice los resultados obtenidos utilizando:

- i. Función de activación 'sigmoid' y función de costo 'ECM' (error cuadrático medio)
- ii. Función de activación 'sigmoid' y función de costo 'EC\_binaria' (entropía cruzada binaria)
- iii. Función de activación 'tanh' y función de costo 'ECM' (error cuadrático medio)

## Regresión Logística multiclase

### Ejercicio 12

Utilice una red neuronal formada por una única capa de salida de 3 neuronas para clasificar las muestras de vino del archivo Vinos.csv descrito en el ejercicio anterior.

Realice 30 ejecuciones independientes utilizando el 60% y 80% de los ejemplos como entrenamiento y el resto como testeo.

Utilice un máximo de 400 iteraciones y velocidades de aprendizaje 0.1, 0.2 y 0.3.

Complete la siguiente tabla con los resultados de las siguientes configuraciones

- a) Función de activación 'sigmoid' y función de costo 'ECM' (error cuadrático medio)
- b) Función de activación 'sigmoid' y función de costo 'EC\_binaria' (entropía cruzada binaria)
- c) Función de activación 'tanh' y función de costo 'ECM' (error cuadrático medio)
- d) Capa 'Softmax' y función de costo 'EC' (entropía cruzada)

Train-test	Función y Costo	alfa	Iteraciones	Accuracy train	Veces (datos Train)	Accuracy test	Veces (Datos Test)
80-20	'sigmoid' ECM	0,3					
		0,2					
		0,1					
	'sigmoid' EC_bin	0,3					
		0,2					
		0,1					
	tanh	0,3					
		0,2					
		0,1					
	'Softmax' EC	0,3					
		0,2					
		0,1					
60-40	'sigmoid' ECM	0,3					
		0,2					
		0,1					
	'sigmoid' EC_bin	0,3					
		0,2					
		0,1					
	tanh	0,3					
		0,2					
		0,1					
	'Softmax' EC	0,3					
		0,2					
		0,1					