



IPN-UPIITA

## Redes Neuronales

Reporte R04

Dr. Rafael Martínez Martínez

Academia de sistemas

ramartinezr@ipn.mx

---

### Instrucciones:

- Cada problema/ejercicio debe tener procedimiento ordenado y completo que justifique adecuadamente la respuesta anotada.
  - Si falta el procedimiento o este no justifica la respuesta anotada entonces el problema vale 0 puntos aunque la respuesta sea correcta.
- 

## Contenido

### Problema 1 (100 puntos)

1

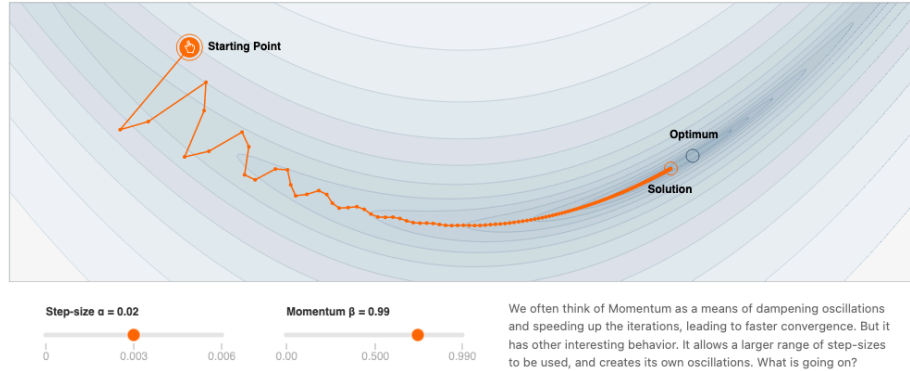
### Problema 1 (100 puntos)

La Figura 1 fue tomada del artículo: [Why Momentum Really Works](#). La Figura 1 muestra la trayectoria de la sucesión de búsqueda del mínimo para una función cuadrática (no especificada) y un punto inicial (interactivo). En particular la trayectoria es la generada por el algoritmo de **momentum** para sus parámetros<sup>1</sup>  $\alpha = 0.003$  (Step-size) y  $\beta = 0.75$  (momentum). La entrega consiste en responder preguntas respecto a la primer parte del artículo<sup>2</sup>. Lo ideal es que dieras una primer lectura a todo la información y después contestaras regresando a la información específica de la pregunta.

---

<sup>1</sup>Valores aproximados, pues hay un error en la visualización inicial que indican  $\alpha = 0.02$ ,  $\beta = 0.99$ , si se intenta probar con los valores corregidos se aproxima a la Figura mostrada inicialmente en la aplicación

<sup>2</sup>La segunda prte del artículo se entregará en el siguiente reporte



**Figura 1.** Trayectoria del momentum

- i. Sin mover el punto de inicio. Reporta gráficas para las Figuras generadas por al aplicacion para los siguientes casos

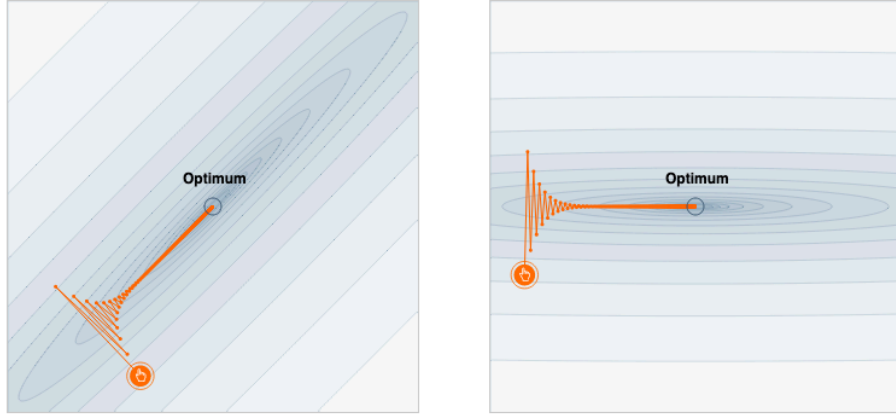
- $\alpha = 0.003$  y  $\beta = 0$
- $\alpha = 0.003$  y  $\beta = 0.99$
- $\alpha = 0.0035$  y  $\beta = 0.75$
- $\alpha = 0.0045$  y  $\beta = 0.75$
- $\alpha = 0.0055$  y  $\beta = 0.75$

¿Cuáles son tus primeras impresiones al respecto?

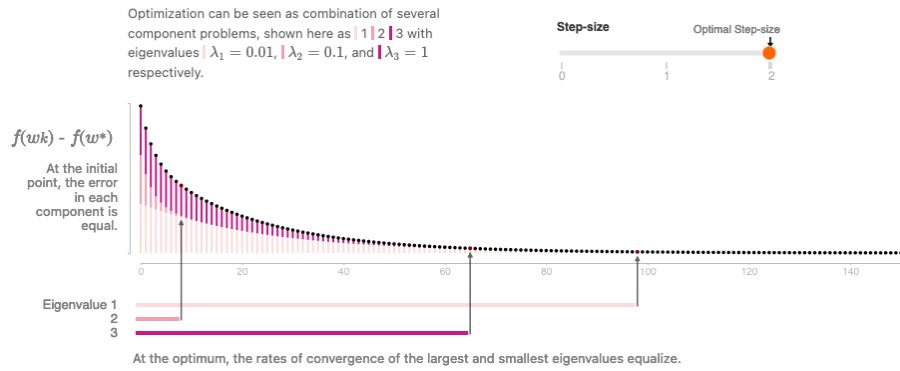
- En los algoritmos vistos en clase hablamos de la *tasa de aprendizaje* ¿Cuál es el nombre que recibe este parámetro en el artículo?
- Anotas las ecuaciones que se refieren al algoritmo de momentum. En este punto debería tenerse claro en donde intervienen los parámetros  $\alpha$  y  $\beta$  antes utilizados.
- ¿Qué sucede si en la Figura 2 el punto de inicio está en la línea de acción del vector propio? Reporta la figura que ilustre esto.
- Para el algoritmo de gradiente descendiente puede caracterizarse el vector de error<sup>3</sup> en cada iteración  $k$  en términos de las componentes de la condición inicial,  $\alpha$ , los valores propios y los vectores propios (se a supuesto una forma cuadrática), es decir

$$w^k - w^* = Qx^k = \sum_i^n x_i^0 (1 - \alpha\lambda_i)^k q_i \quad (1)$$

<sup>3</sup>La diferencia entre la aproximación  $w^k$  (en el paso  $k$ ) y el mínimo  $w^*$



**Figura 2.** Trayectoria de gradiente descendiente



**Figura 3.** Geometría del error de evaluación

de forma similar, se puede caracterizar al error entre la evaluación de la función en la iteración  $k$  y la evaluación en el mínimo

$$f(w^k) - f(w^*) = \sum (1 - \alpha\lambda_i)^{2k} \lambda_i [x_i^0]^2 \quad (2)$$

Explica la Figura 3 a detalle (todos los elementos gráficos son la ayuda). Puede servir mover los valores del Step-size para entender mejor la gráfica

- vi. ¿Por qué  $|1 - \alpha\lambda_i| < 1$  garantiza convergencia?
- vii. Explica que es una *tasa*. Hay muchos contextos para ello, en particular reporta que es la tasa de convergencia de una sucesión (podrías encontrarlo como tasa de convergencia de un algoritmo). La idea es tener claro lo que sigue

- viii. Reporta el valor óptimo de  $\alpha$  y el valor óptimo de la tasa. Para concluir esta sección se debe tener en mente que cuando  $\kappa = \lambda_n/\lambda_1 \gg 1$  el algoritmo de gradiente descendiente tiene comportamientos no deseados.
- ix. Debido al background que tenemos la sección: *Example: Polynomial Regression*, puede ser difícil de leer, se proporcionan algunos elementos para que lectura sea más digerible. La sección es un ejemplo diferente para ilustrar un fenómeno que se desprende de lo abordado en el pregunta v.

Si tenemos los polinomios  $\{p_1(\xi) = 1, p_2(\xi) = \xi, p_3(\xi) = \xi^2\}$ , el vector  $a(\xi) = -1 + 2\xi - 3\xi^2$ , puede ser expresado por el vector de coordenadas  $\bar{a} = [-1 \ 2 \ -3]^T$  en  $\mathbb{R}^3$  (como espacios vectoriales son isomorfos). Así se puede hacer la asociación de bases  $p_1(\xi) \rightarrow [1, 0, 0]^T$ ,  $p_2(\xi) \rightarrow [0, 1, 0]^T$  y  $p_3(\xi) \rightarrow [0, 0, 1]^T$ . Así un problema que involucre propiedades de espacio vectorial en el espacio de polinomios, puede trasladarse a un problema en  $\mathbb{R}^n$ , donde  $n$  depende del grado del polinomio

Supongamos que tenemos las siguientes mediciones

$\xi_i$ (las $x$ )	$d_i$ (las $y$ )
-3	-0.95
-2	0.5
-1	0.7

Supongamos que queremos encontrar la *mejor combinación lineal* dada por

$$\text{model}(\xi) = w_1 p_1(\xi) + w_2 p_2(\xi) + w_3 p_3(\xi) \quad p_i = \xi \mapsto \xi^{i-1}$$

De tal forma que los  $w_i$  minimicen la siguiente expresión, esa expresión hace referencia a la suma de los errores al cuadrado entre el valor observado  $d_i$  y el calculado con el modelo propuesto (en este sentido será la mejor combinación lineal)

$$\min_w \frac{1}{2} \sum (\text{model}(\xi_i) - d_i)^2 = \frac{1}{2} \|Zw - d\|^2 = \frac{1}{2} (Zw - d)^T (Zw - d)$$

Con ayuda de la sección correspondiente del artículo y lo anterior. Escribe  $Z$ ,  $w$  y  $d$  para el ejemplo antes planteado. A este punto debe de quedar clara la equivalencia entre la suma y la forma matricial.

La sección va en esta dirección, se busca encontrar  $w$  que minimice la expresión del error. Así, si se hace un cambio de base o cualquier transformación de esta en  $\mathbb{R}^n$  se verá reflejado en la correspondiente base de polinomios y viceversa. Observemos que queremos minimizar una forma cuadrática en  $w$ .

- x. Explica las Figuras 4, 5 y 6. Resume las conclusiones de la sección correspondiente. La mejor forma de entender es manipular los elementos gráficos de las aplicaciones. Además recuerda que siempre puedes buscar información adicional.

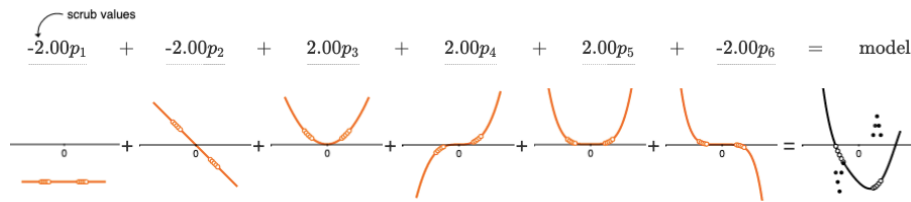


Figura 4. Combinación lineal base c  nica

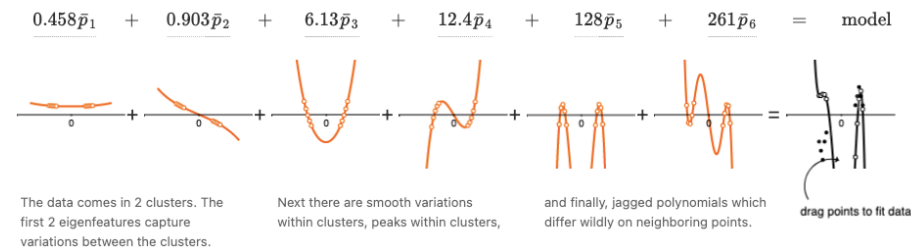


Figura 5. Combiaci  n lineal eigenbase

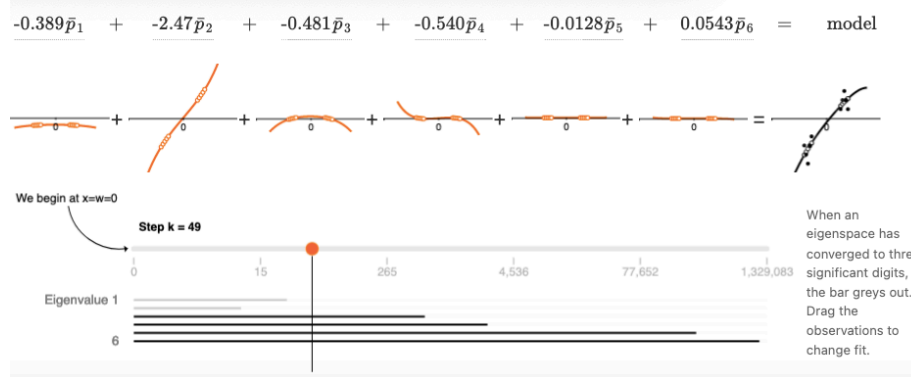


Figura 6. Combiaci  n lineal eigenbase

