

# Resumen del Artículo

## Comparing the moore-penrose y el descenso de Gradeinte

### Describir

- La regresión lineal es uno de los medios más clásicos en estadísticas y áreas durante algemático, busca ayudar con modelo lineal y base automático el error cuadrático (mínimos cuadrados o OLS)
- Una vez formulado se problema existen puntualmente dos entonces a resolverlo:
  1. Solución Analítica mediante la Posición media de mópre-ferrose.
  2. Atelado Técnico como el gradiente desengueñe.

Cada enfoque tiene ventajas y desventajas. Complejidad estándardamente estandarizada, el Artículo tiene como objetivo comparar estos modelos tanto decorrendo como con experimentos, para demorar en qué situaciones uno es preferible sobre el otro.

---

### Describe:

1. Para que condiciones (dimensionalidad, lanzón de muestra, conductor del sistema etc.), la posición interna es más eficiente o más estable y cuando conviene usar la gradiente desengueñe?
- 

### Formulación de la Regresión Lineal

- Se tienen datos CX-y con  $x \in R^{n+1}$  en muestras de características  $y \in R^n$ .
- Se modela  $y = xB + e_y$  con el medio medio cero.
- El objetivo es minimizar la función perátna.

$$L(B) = \|xB - y\|^2 = (xB - y)^T(xB - y).$$

como  $L(B)$  es convexa, cualquier mínimo local tambien es global.

### Solución con la Posición interna

El método clásico abajo es:

$$x^T x B = x^T y \quad y \quad as^t obtener \hat{\beta}_{p,m,l} = (x^T x)^{-1} x^T y,$$

cuando  $x$  es invertido: en general, si la misma  $x$  no es de rango completo se usa la posición interna de mópre-ferrose  $x^T$ ,  $\hat{\beta}_{p,m,l} = x^T y$ .

### Formulación

- **Formulación**
  - una unas operación: compañía alímpole colosa para matrices grandes
  - una inversión/descon frecuencia puntería si  $x^T x$  es mal condicionada.

### Resultados exactos

- **Resultados exactos**
  - las escala bajo cuando el número de muestras o características no iterativos?

## Gradiente desenciente aplicado a regresión lineal

Sé minimiza iterativamente $V_L(\beta) = 2x^T(2\beta - y)$	$\hat{\beta}_{k,l} = \hat{\beta}_k - n\nabla(L\beta_k)$ donde $n$ es la taza de aprendizaje - el modelo requiere una inicialización elegir criterio de pasado (número máximo de variaciones) no da solución, evadir en un solo paso, sino convergir rápidamente al mismo - convergir rápidamente al mismo - convergir tanto si la tasa de amenazaje es pequeña - sensible a la conducta puntería - debe elegirse cuidadosamente el paso y criterios
$\hat{\beta}_{k,l} = \hat{\beta}_k - 2nx^T(x\beta_k - y)$ Ventaja - escalable no necesita invertir matrices enormes - efectivo puede adaptarse a datasets (gradiente, ejecutivo)	

## Emisión

### ■ Emisión

- Elemento normal: un ácido superficie de gestión matrices  $x$  en distintos temas. Es importante que ambos métodos tiempo de computa-error de predicción están India numerice.

## Plan profesional respecto y medianos con matrices de productividad

### ■ Plan profesional respecto

- por dicho muy grande el gracioso tipo de productividad, la posición interna ofrece una solución rápida.

## Comparación de rendimiento entre el modelo lineal y el mejo Gradiente Desendente.

Objetivo evaluar la eficiencia, computacional del método de regresión lineal im\_model frente al algoritmo de Gradiente Desendente en términos de tiempo de ejecución (Segundos) y uso de memoria (byles) para diferentes trampas de muestras.

**1. Resultados fabulares: a composición se muestran los resultados del rendimiento con los valores de tiempo.**

Tamaños (N)	Método	Tiempo (S)	Memoria (byles)
100	Im_model	O	64,680
400	Gradiente	O	4,830,776
4000	Im_model	O	619,048
4000	Gradiente	O	17,649,200
40000	Im_model	O	6213,928
40000	Gradiente	O	176,049,200
400000	Im_model	O	64,881,432
400000	Gradiente	O	1,760,019,200
500,000	Im_model	O	308,197,160
500,000	Gradiente	O	8,800,019,200
1,000,000	Im_model	I	646,394,464
1,000,000	Gradiente	I7	17,678,009,744
2M	Im_model	2	1,232,780,072
2M	Gradiente	3M	35,200,039,084

analisis: los resultados que claramente se evidencian una ventaja clara del método im\_model frente al gradiente Desendente.

- Velocidad CTempo: el im\_model está inscrito mejor para un descargamiento al gradiente Desendente requiere su segundo tipo (lo hace aproximadamente frueces más lentos que im\_model (2 segundos).
- Uso de recursos Chemo(ia) (a diferencia de abrumadora para la muestra grande, el gradiente consume más de 35 millones de byles mientras que im\_model se tumba aproximadamente 12 millones de byles demostrando una eficiencia de recursos significadamente superior.

La consecuencia para las líneas de regresión lineal con volúmenes de dolor mediante girables, el im\_model ofrece un rendimiento superior y un rendimiento superior y un uso de más recursos más eficientes que el gradiente Desendente.

Y	propios CSI	Memory	Time
2	modelo_im L—Im (y-x)	198.4	380
3	gradiente_desc (x,y)	-20638.74	28046.8
4	3		

---

Classic