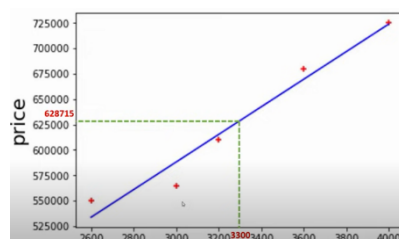


1 MACHINE LEARNING – REGRESIÓN LINEAL

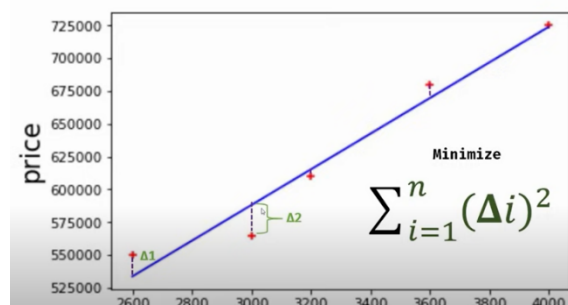
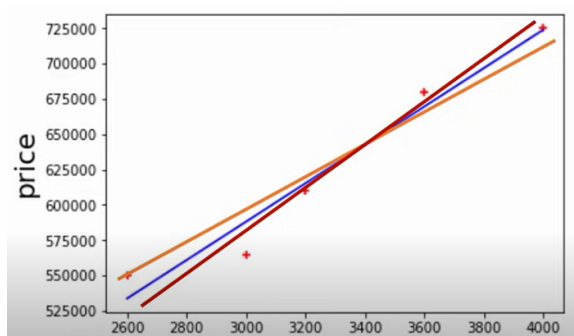
El problema planteado es el siguiente: Tenemos un listado de casas con información básica y queremos predecir el precio de estas según el área.

area	price
2600	550000
3000	565000
3200	610000
3600	680000
4000	725000

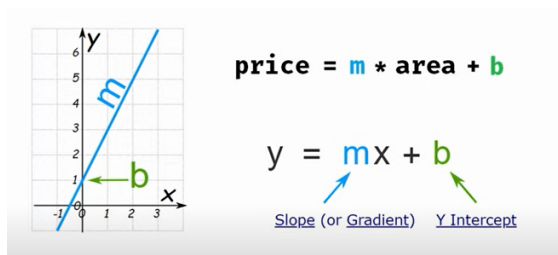
Si representamos los datos con un SCATTERPLOT, obtenemos la siguiente gráfica en la que podemos calcular de manera aproximada los precios de cualquier casa siguiendo este modelo lineal.



Pero ... qué línea trazamos? Matemáticamente, la que minimice la suma de las distancias entre los puntos y la línea.



Debemos saber que una ecuación lineal (recta) viene definida por:



m: Pendiente de la recta

b: Ordenada en el origen (altura que tiene la recta cuando $x=0$)

En Python lo llamamos **coef_ e intercept_**



GENERALITAT
VALENCIANA



UNIÓN EUROPEA
Fondo Social Europeo
El FSE invierte en tu futuro

03008915 C/ Ferrocarril, 22, 03570 La Vila Joiosa Tel 966870140 Fax 966870141 <http://portal.edu.gva.es/iesmarcoszaragoza>

En el ejemplo, price es la variable dependiente y area es la variable independiente. Para la aplicación de algoritmos de machine learning vamos a usar la librería **Scikit-learn (sklearn)**

Dentro de esta librería, usaremos el modelo lineal:

```
X = df.drop('price',axis=1)
from sklearn import linear_model
reg = linear_model.LinearRegression()
reg.fit(X,df.price)
```

Con este método (fit) lo que hacemos es entrenar al algoritmo con los valores de x (área) e y (precio)

podemos predecir algún valor con:

```
reg.predict([[3000]])
```

El modelo lineal tiene los coeficientes de la línea de regresión que hemos hablado anteriormente. Para conocerlos, podemos acceder a:

```
reg.coef_ y reg.intercept_
```

Podemos leer un csv de áreas que queremos predecir:

```
d = pd.read_csv('areas.csv')
p = reg.predict(d)
```