

# regression

November 11, 2024

```
[1]: import numpy as np
import pandas as pd
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

import matplotlib.pyplot as plt
```

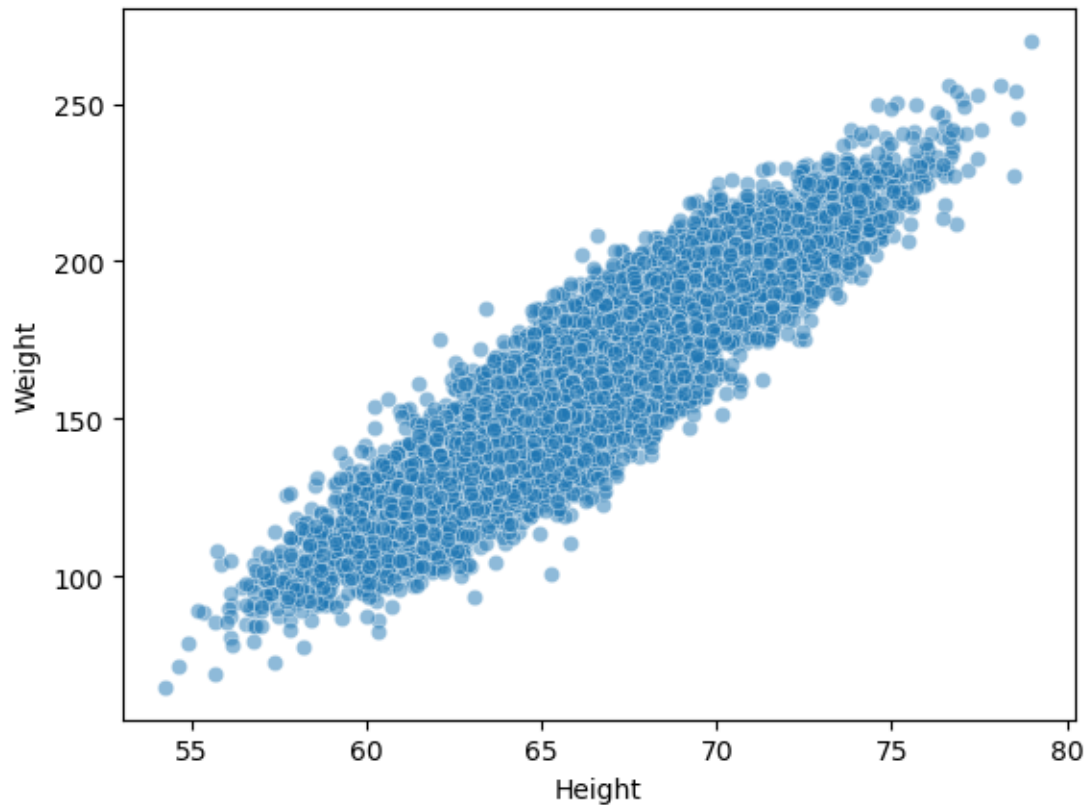
```
[2]: url = "https://raw.githubusercontent.com/zerotodeeplearning/ztdl-masterclasses/
↳master/data/"
```

```
[4]: df = pd.read_csv(url + 'weight-height.csv')
```

```
[5]: df.head()
```

```
[5]:   Gender      Height      Weight
0   Male  73.847017  241.893563
1   Male  68.781904  162.310473
2   Male  74.110105  212.740856
3   Male  71.730978  220.042470
4   Male  69.881796  206.349801
```

```
[6]: sns.scatterplot(data=df, x='Height', y='Weight', alpha=0.5);
```



```
[7]: X = df[['Height']].values  
     y = df['Weight'].values
```

```
[8]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,  
     ↪ random_state=0)
```

```
[9]: model = LinearRegression()
```

```
[10]: model.fit(X_train, y_train)
```

```
[10]: LinearRegression()
```

```
[11]: model.score(X_train, y_train)
```

```
[11]: 0.8544074009765417
```

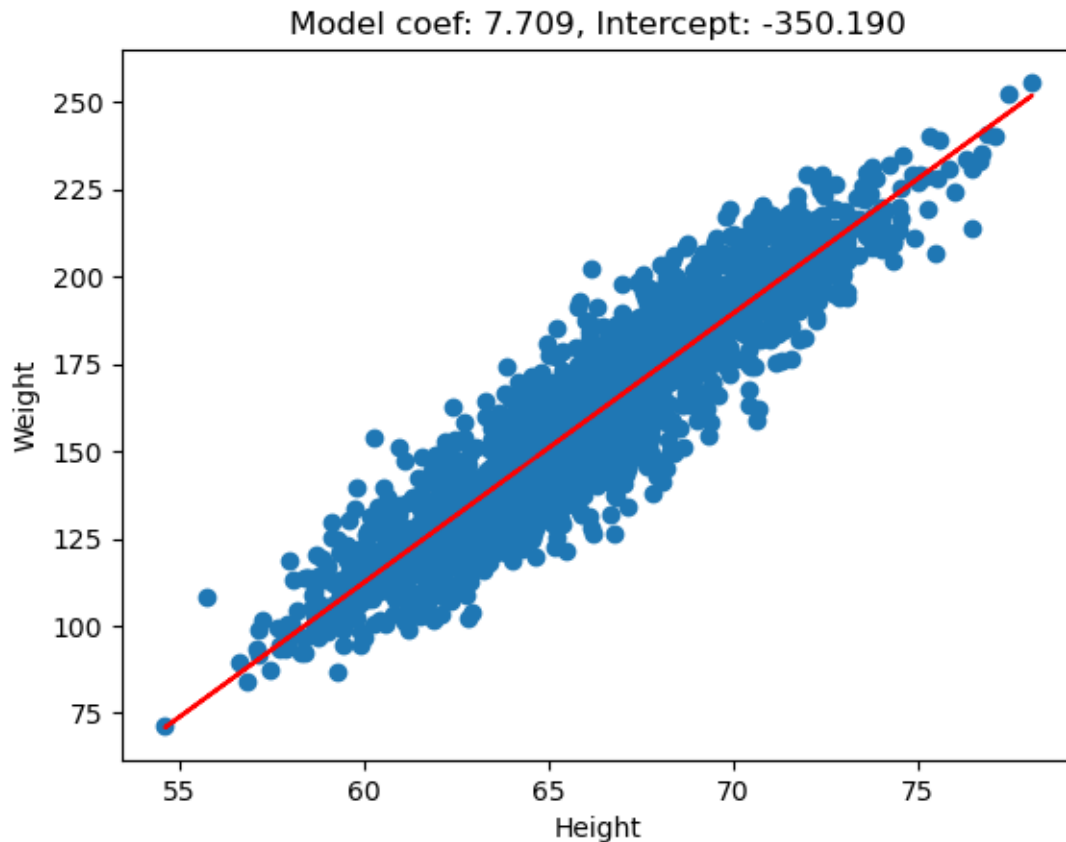
```
[13]: model.score(X_test, y_test)
```

```
[13]: 0.8581952493935256
```

```
[12]: y_pred_test = model.predict(X_test)
```

```
[14]: plt.scatter(X_test, y_test)
plt.plot(X_test, y_pred_test, color='red')
plt.title("Model coef: {:.3f}, Intercept: {:.3f}".format(model.coef_[0],
↳model.intercept_))
plt.xlabel("Height")
plt.ylabel("Weight")
```

```
[14]: Text(0, 0.5, 'Weight')
```

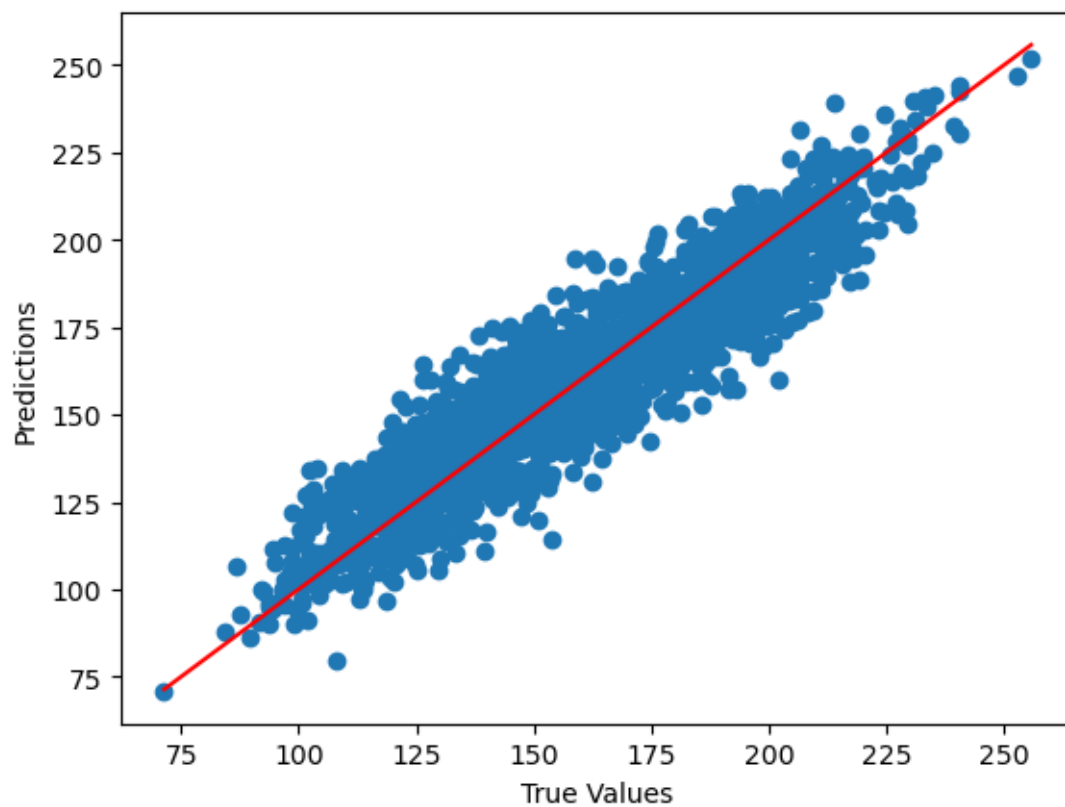


### 0.0.1 Compare true and predicted values ( $y$ vs $\hat{y}$ )

```
[21]: plt.scatter(y_test, y_pred_test)
plt.xlabel("True Values")
plt.ylabel("Predictions")

m = y_test.min()
M = y_test.max()

plt.plot((m, M), (m, M), color='red');
```



[ ]: