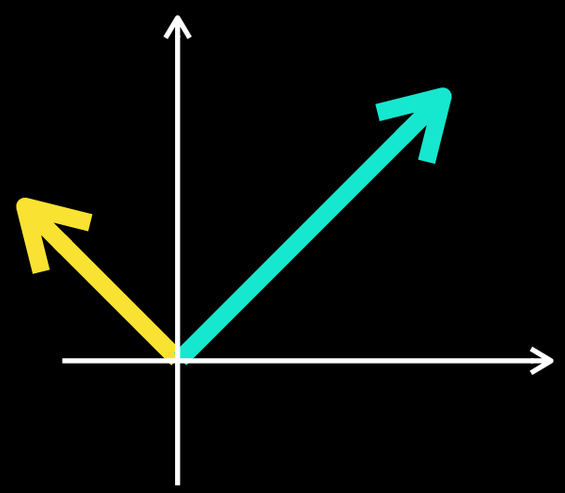
Linear regression: Californian houses

Linear Algebra Essentials



Example: Californian house prices

```
import pandas as pd
```

```
data = pd.read_csv("housing.csv")
data = data.dropna()[data.columns[0:9]]

data.head()
```

= X

= y

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value
0	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252	452600.0
1	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	358500.0
2	-122.24	37.85	52.0	1467.0	190.0	496.0	177.0	7.2574	352100.0
3	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431	341300.0
4	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462	342200.0

```
X = data[data.columns[0:8]].values.T
X = np.vstack([np.ones(X.shape[1]), X])
X[:, :3]
```

```
array([[ 1.0000e+00,  1.0000e+00,  1.0000e+00],
       [-1.2223e+02, -1.2222e+02, -1.2224e+02],
       [ 3.7880e+01,  3.7860e+01,  3.7850e+01],
       [ 4.1000e+01,  2.1000e+01,  5.2000e+01],
       [ 8.8000e+02,  7.0990e+03,  1.4670e+03],
       [ 1.2900e+02,  1.1060e+03,  1.9000e+02],
       [ 3.2200e+02,  2.4010e+03,  4.9600e+02],
       [ 1.2600e+02,  1.1380e+03,  1.7700e+02],
       [ 8.3252e+00,  8.3014e+00,  7.2574e+00]])
```

```
y =  data[data.columns[8]].values
y
```

```
array([452600., 358500., 352100., ...,  92300.,  84700.,  89400.])
```

```
a =  np.linalg.inv(X.dot(X.T)).dot(X).dot(y)
a
```

intercept *coefs*

```
array([-3.58539575e+06, -4.27301205e+04, -4.25097369e+04,  1.15790031e+03,
       -8.24972507e+00,  1.13820707e+02, -3.83855780e+01,  4.77013513e+01,
        4.02975217e+04])
```

```
dv = X.std(axis=1)
dv
```

```
array([0.00000000e+00, 2.00352886e+00, 2.13629539e+00, 1.25914971e+01,
       2.18521609e+03, 4.21374759e+02, 1.13318076e+03, 3.82289871e+02,
       1.89924477e+00])
```

```
dv * a
```

longitude *latitude* *age*

```
array([    -0.        , -85611.02961399, -90813.35501124,  14579.6983295 ,
       -18027.43197782,  47961.17298572, -43497.79848984,  18235.74343777,
        76534.85746792])
```

area *income* *bedroom area* *population* *households*

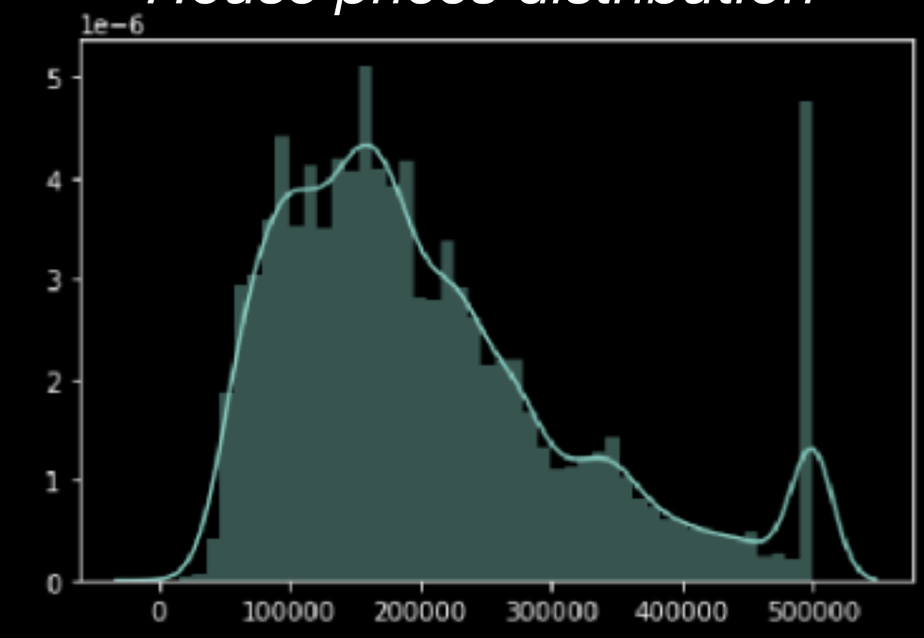
```
(X.T.dot(a) - y).mean()
```

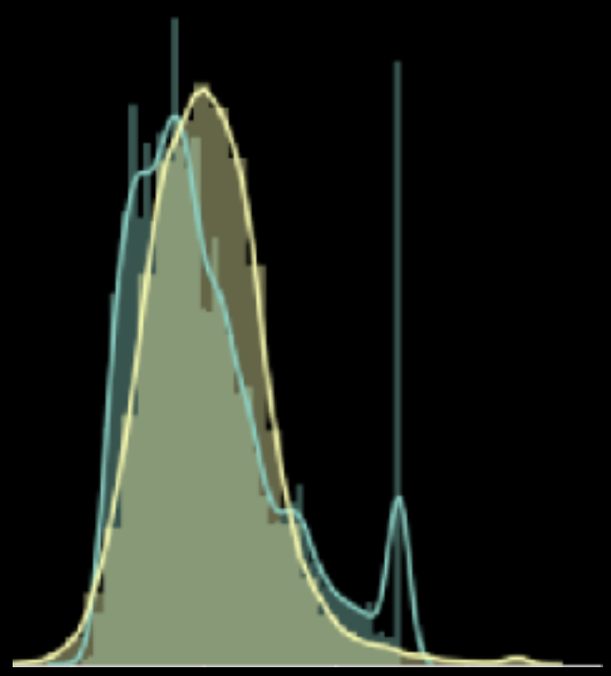
```
-2.2030123056180l5e-06
```

```
(X.T.dot(a) - y).std()
```

```
69556.14839566677
```

House prices distribution





actual vs. predicted distribution