Vectors as Features

- β for each feature

  Now suppose that our feature is a vector of length k

  We have data as before: $(\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2), \ldots, (\mathbf{X}_n, Y_n)$, where

  $$\mathbf{X}_i = (X_{i1}, \ldots, X_{ik})$$

  The linear regression model is now:

  $$Y_i = \sum_{j=1}^{k} \beta_j X_{ij} + \epsilon_i, \text{ where } E(\epsilon_i | \mathbf{X}_i) = 0$$

  To include intercept in the model we can set $X_{i1} = 1$

- Then each $\mathbf{X}_i$ is a vector of length $k + 1$

- as matrix

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

–

- Let $\hat{\boldsymbol{\beta}}$ denote the estimate for $\boldsymbol{\beta}$
- The predicted values are:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

- The errors are given by:

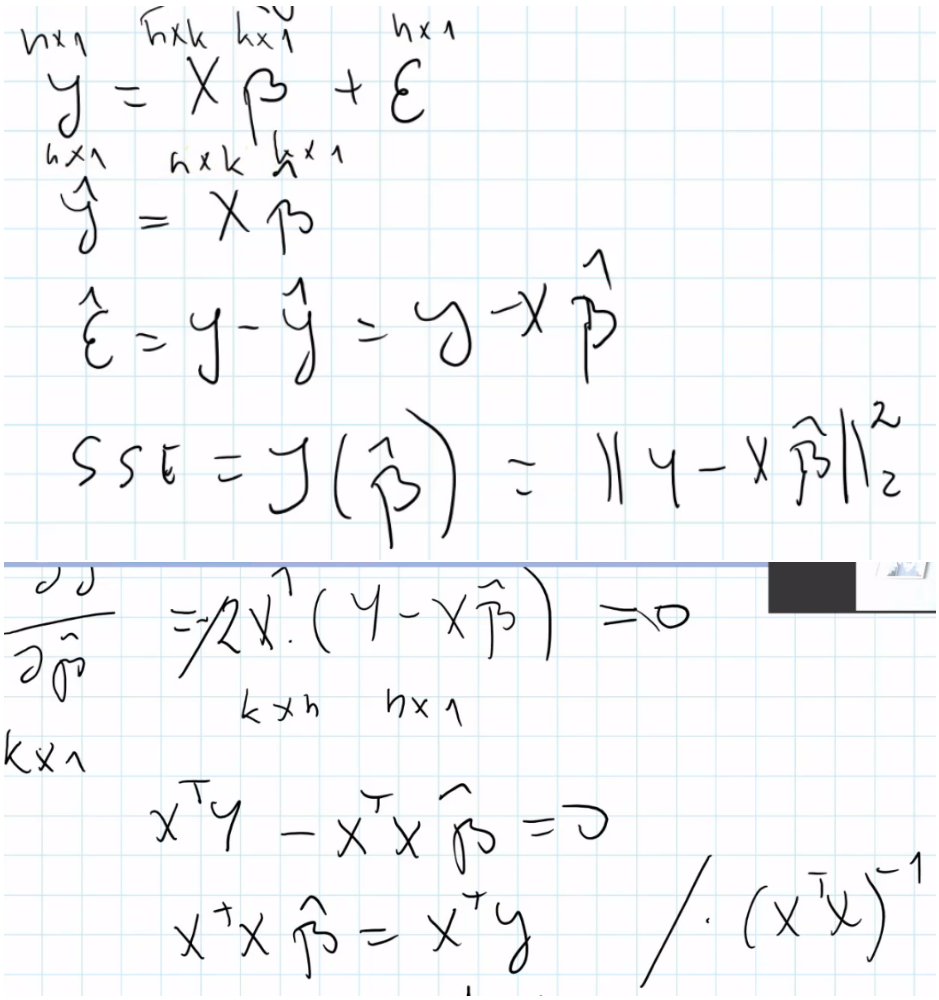
$$\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

- The total sum of squared errors is given by the vector norm:

$$SSE = ||\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}||_2^2$$

Least Squares +

Esimate $\hat{\beta}$

- 

- normal equations
  - $\hat{\beta} = (X^T X)^{-1} X^T y$

Example

We have a video games dataset containing games release dates, price, sales, average and median playtime from Steam, critics rating, and the user rating from from Metacritic reviewing site. We are interested in relationship between user rating and other features. We fit the model and obtain the following results:

Feature	$\hat{\beta}_j$	conf. intervals
Intercept	7.04948665	(6.999040, 7.102739)
Release Date	0.09082842	(0.032808, 0.145047)
Price	-0.1283695	(-0.197546, -0.057582)
Sales	0.03249976	(-0.015483, 0.085458)
Avg. Playtime	0.02901469	(-0.041831, 0.111011)
Md. Playtime	0.01683537	(-0.077188, 0.071164)
Critics Score	0.70902049	(0.647059, 0.775075)