

Note that in building our kNN model for prediction, we did not compute a closed form for \hat{F} . kNN is a **non-parametric** model.

What if we ask the question, "how much more sales do we expect if we double the TV advertising budget?"

We can build a model by first assuming a simple form of f:

$$f(x) = \beta_0 + \beta_1 X$$

where $oldsymbol{eta_1}$ represents the slope and $oldsymbol{eta_0}$ represents the intercept. It then follows that our estimate is:

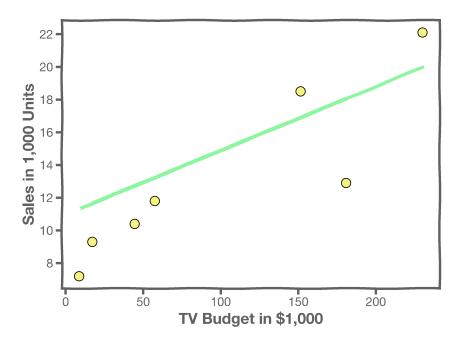
$$\hat{Y} = \hat{f}(x) = \hat{\beta}_0 + \hat{\beta}_1 X$$

where $\hat{m{eta}}_1$ and $\hat{m{eta}}_0$ are estimates of $m{eta}_1$ and $m{eta}_0$, respectively, which we compute using observations.

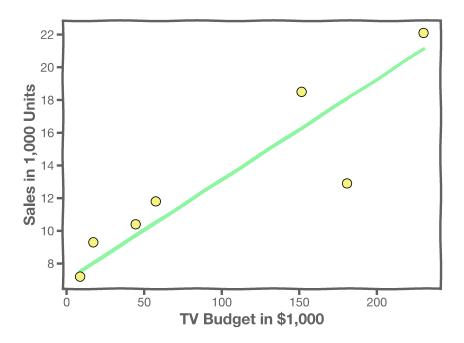
Estimate of the regression coefficients

For a given data set we can draw different lines through the data, we need to determine which line is the best fit.

Is this it?

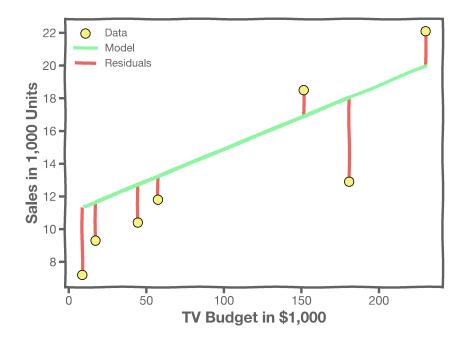


Or this one?



Question: Which line is the best?

To decide which is the best line, we can do the same as we did with the kNN model. We estimate the error for every model by looking at the residuals, the difference between the exact value of \boldsymbol{y} and the predicted \boldsymbol{y} . As shown in the plot below, the regression or model line, predicted sales, is the slanted line in green and the residuals to the exact values of sales are the vertical red lines.



As before, for each observation (x_n, y_n) , the absolute residuals, $r_i = |y_i - \hat{y_i}|$ quantifies the error at each observation.

Again, we aggregate and use the Mean Squared Error, MSE, as our loss function.

$$L\left(eta_0,eta_1
ight) = rac{1}{n}\sum_{i=1}^n \left(y_i - \hat{y_i}
ight)^2$$

The difference from the kNN model is that there is a specific formula for the $\hat{\pmb{y}}$, which is the linear model.

We choose $m{eta_0}$ and $m{eta_1}$ that minimize the predictive errors made by our model, that is, minimize our loss function.

Then the optimal values, $\hat{oldsymbol{eta}}_{oldsymbol{0}}$ and $\hat{oldsymbol{eta}}_{oldsymbol{1}}$, should be:

$$\hat{eta}_0,\hat{eta}_1=rg\min_{eta_0,eta_1}L(eta_0,eta_1)$$

We call this "fitting" or "training" the model.

Discussion Board (External resource)

Click OK to have your username and e-mail address sent to a 3rd party application.

| ⟨ Previous N | lext > |
|--------------|--------|
|--------------|--------|

© All Rights Reserved





Legal

Terms of Service & Honor Code

Privacy Policy

Accessibility Policy

Trademark Policy

<u>Sitemap</u>

Cookie Policy

Your Privacy Choices

Connect

<u>Blog</u>

Contact Us

Help Center

Security

Media Kit















© 2023 edX LLC. All rights reserved.

深圳市恒宇博科技有限公司 <u>粤ICP备17044299号-2</u>