## Least squares derivation of linear regression

$$Lossfunction = \sum_{i=1}^{n}(y_i - \hat{y_i})^2$$

Define loss function and expand:

$$Q = \sum_{i=1}^{n}(y_i - \hat{y_i})^2, \hat{y_i} = mx_i + b$$

$$Q(m,b) = \sum_{i=1}^{n}(y_i - (mx_i + b))^2$$

$$Q(m,b) = \sum_{i=1}^{n}(y_i - mx_i - b)^2$$

$$Q(m,b) = \sum_{i=1}^{n}(y_i^2 - y_imx_i - y_ib - y_imx_i + m^2x_i^2 + mx_ib - y_ib + mx_ib + b^2)$$

$$Q(m,b) = \sum_{i=1}^{n} y_i^2 - \sum_{i=1}^{n} y_imx_i - \sum_{i=1}^{n} y_ib - \sum_{i=1}^{n} y_imx_i + \sum_{i=1}^{n} m^2x_i^2 + \sum_{i=1}^{n} mx_ib - \sum_{i=1}^{n} y_ib + \sum_{i=1}^{n} mx_ib + \sum_{i=1}^{n} b^2$$

$$Q(m,b) = \sum_{i=1}^{n} y_i^2 + \sum_{i=1}^{n} m^2x_i^2 + \sum_{i=1}^{n} b^2 - \sum_{i=1}^{n} 2y_imx_i - \sum_{i=1}^{n} 2y_ib + \sum_{i=1}^{n} 2mx_ib$$

Differentiate with respect to **m** and **b**:

$$\frac{\partial Q}{\partial m} = 2m\sum x_i^2 - 2\sum_{i=1}^{n} y_ix_i + 2b\sum_{i=1}^{n} x_i$$

$$\frac{\partial Q}{\partial b} = 2bN - 2\sum y_i + 2m\sum x_i, N = number\ of\ points$$

We want to minimize m and b, so set both to 0:

$$2m\sum x_i^2 - 2\sum y_ix_i + 2b\sum x_i = 0$$

$$2bN - 2\sum y_i + 2m\sum x_i = 0$$

Solve for b first:

$$b = \frac{2\sum y_i - 2m\sum x_i}{2N}$$

$$b = \frac{\sum y_i - m\sum x_i}{N}$$

Substitute b and solve for m:

$$2m\sum x_i^2 - 2\sum y_ix_i + 2\frac{\sum y_i - m\sum x_i}{N}\sum x_i = 0$$

$$m \sum x_i^2 - \sum y_i x_i + \frac{\sum y_i - m \sum x_i}{N} \sum x_i = 0$$

$$m \sum x_i^2 - \sum y_i x_i + \frac{\sum y_i \sum x_i - m(\sum x_i)^2}{N} = 0$$

$$mN \sum x_i^2 - N \sum y_i x_i + \sum y_i \sum x_i - m(\sum x_i)^2 = 0$$

$$mN \sum x_i^2 - m(\sum x_i)^2 = N \sum y_i x_i - \sum y_i \sum x_i$$

$$m(N \sum x_i^2 - (\sum x_i)^2) = N \sum y_i x_i - \sum y_i \sum x_i$$

$$m = \frac{N \sum y_i x_i - \sum y_i \sum x_i}{N \sum x_i^2 - (\sum x_i)^2}$$

This solution assumes:

1. Normally distributed residuals.

## Least squares derivation of linear regression (abridged)

1. Define loss function and expand:

$$Q(m, b) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - (mx_i + b))^2$$

...

$$Q(m, b) = \sum_{i=1}^{n} y_i^2 + \sum_{i=1}^{n} m^2 x_i^2 + \sum_{i=1}^{n} b^2 - \sum_{i=1}^{n} 2y_i m x_i - \sum_{i=1}^{n} 2y_i b + \sum_{i=1}^{n} 2m x_i b$$

2. Differentiate with respect to **m** and **b** and set to 0 because we want to minimize error:

$$\frac{\partial Q}{\partial m} = 0 = 2m \sum x_i^2 - 2 \sum_{i=1}^{n} y_i x_i + 2b \sum_{i=1}^{n} x_i$$

$$\frac{\partial Q}{\partial b} = 0 = 2bN - 2 \sum y_i + 2m \sum x_i$$

$$N = number\ of\ points$$

3. Solve for **m** and **b**:

$$m = \frac{N \sum y_i x_i - \sum y_i \sum x_i}{N \sum x_i^2 - (\sum x_i)^2}$$

$$b = \frac{\sum y_i - m \sum x_i}{N}$$

## Maximum likelihood derivation of linear regression (abridged)

1. PDF for normally-distributed variable: $X \sim \mathcal{N}(\mu, \sigma^2)$

   $$PDF(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

2. PDF, with linear equation in place of the mean.

   $$P(y|x; m, b, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-(mx+b))^2}{2\sigma^2}}$$

3. Likelihood function for all observed points (x, y) is the product of the probability density for each point:

   $$L(m, b, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \prod_{i=1}^{n} e^{\frac{-(y_i-(mx_i+b))^2}{2\sigma^2}}$$

4. Goal: find parameters **b**, **m** and $\sigma$ that **maximize *L*.**

1. Convert to log likelihood for easier math, **maximize *log(L)***:

   $$log(L) = log[\frac{1}{\sqrt{2\pi\sigma^2}} \prod_{i=1}^{n} e^{\frac{-(y_i-(mx_i+b))^2}{2\sigma^2}}]$$

   ...

   $$log(L) = -\frac{1}{2} log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - (mx_i + b))^2$$

2. Or, **minimize** negative log likelihood:

   $$-log(L) = \frac{1}{2} log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - (mx_i + b))^2$$

3. Let's *imagine* that our variance term is a fixed constant.

   $$-log(L) = \sum_{i=1}^{n} (y_i - (mx_i + b))^2$$