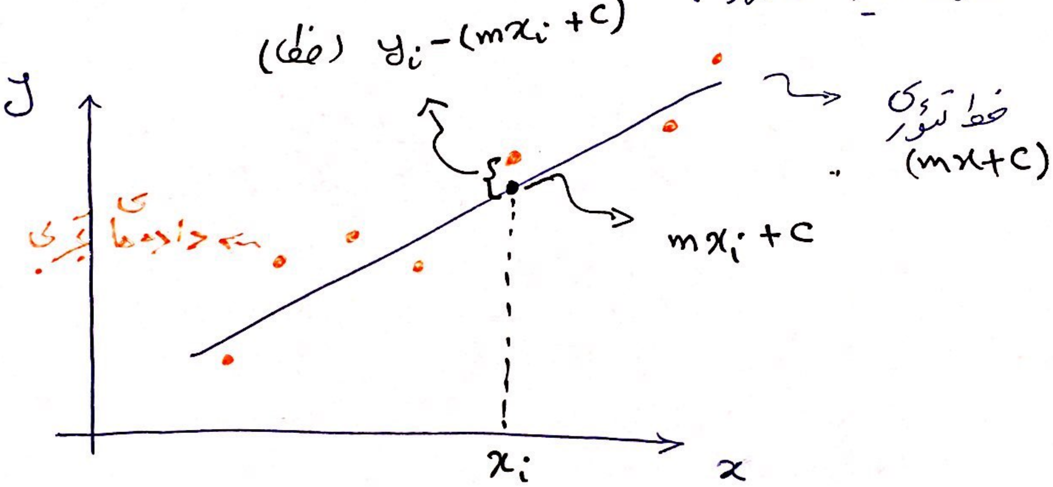
(1)

برازش (Fitting) به روش کمترین مربعات

معمولاً داده‌های تجربی دارای خطا یا اصطلاحاً نویز هستند. بنابراین داده‌های تجربی کاملاً بر روی نموداری که تئوری پیش بینی می‌کند قرار نمی‌گیرند. برای مثال فرض کنید در یک آزمایش داده‌های اندازه‌گیری شده باید رفتار خطی از خود نشان دهند ولی در واقعیت به دلیل وجود نویزها نمودار کاملاً خطی نمی‌شود. نمودار زیر مثالی از داده‌های تجربی است که دارای نویز هستند:



یکی از روش‌هایی که می‌توان بهترین خط یعنی $y=mx+c$ را پیدا کنیم، استفاده از روشی به نام کمترین مربعات است. در این روش برای پیدا کردن بهترین شیب (m) و بهترین عرض از مبدا (c) چنین عمل می‌کنیم:

توان دوم

یک خط فرضی با معادله $y=mx+c$ را در نظر می‌گیریم. فاصله هر نقطه (هر داده) از این خط به صورت $(\underbrace{mx_i+c}_{\text{مولفه } y \text{ بر روی خط}}-y_i)^2$ خواهد شد. جمع کل این مقادیر را با χ^2 نمایش می‌دهیم:

$$\chi^2=\sum_{i=1}^{N}(mx_i+c-y_i)^2$$

مقدار χ^2 را می‌توان به صورت زیر بسط داد:

$$\chi^2=\sum_{i=1}^{N}(mx_i+c-y_i)^2$$

$$=\sum_{i=1}^{N} m^2x_i^2+c^2+y_i^2+2mcx_i-2mx_iy_i-2cy_i$$

قاعده: بهترین خط، خطی است که χ^2 را کمینه کند (کمترین مربعات)، بنابراین از χ^2 نسبت به m و c مشتق می‌گیریم و آن را برابر با صفر قرار می‌دهیم:

$$\frac{\partial^2\chi}{\partial m}=0 \longrightarrow 2m\sum_{i=1}^{N}x_i^2+2c\sum x_i-2\sum x_iy_i=0$$

$$\Rightarrow \boxed{m\sum_{i=1}^{N}x_i^2+c\sum_{i=1}^{N}x_i-\sum_{i=1}^{N}x_iy_i=0} \quad (1)$$

$$\frac{\partial^2\chi}{\partial c}=0 \longrightarrow 2cN+2m\sum x_i-2\sum y_i=0$$

$$\Rightarrow \boxed{m\sum_{i=1}^{N}x_i+cN-\sum_{i=1}^{N}y_i=0} \quad (2)$$

برای راحتی نمادگذاری‌های زیر را انجام می‌دهیم:

$$E_x=\frac{1}{N}\sum_{i=1}^{N}x_i,\quad E_y=\frac{1}{N}\sum_{i=1}^{N}y_i,\quad E_{xx}=\frac{1}{N}\sum_{i=1}^{N}x_i^2,\quad E_{xy}=\frac{1}{N}\sum x_iy_i$$

بنابراین معادلات ① و ② را می‌توان به صورت زیر بازنویسی کرد:

$$\frac{①}{N} = 0 \longrightarrow mE_{xx} + cE_x = E_{xy}$$

$$\frac{②}{N} = 0 \longrightarrow mE_x + c = E_y$$

با حل این دو معادله به صورت همزمان، m و c به صورت زیر بدست می‌آیند:

$$m = \frac{E_{xy} - E_x E_y}{E_{xx} - E_x^2} \quad \text{و} \quad c = \frac{E_{xx} E_y - E_x E_{xy}}{E_{xx} - E_x^2}$$

روش کمترین مربعات، یک روش جامع برای حل چنین مسئله‌ای است.

این روش به صورت ماتریسی در کتابخانه numpy ~~[illegible]~~ پیاده‌سازی شده است.

برای اطلاعات بیشتر می‌توانید به تابع زیر مراجعه کنید:

numpy.linalg.lstsq