Least squares is a special case of an optimization problem. The objective function is the sum of the squared distances. The solution can be found analytically.

Gradient descent is an algorithm to construct the solution of an optimization problem approximately. The benefit is that it can be applied to any objective function, not just squared distances.

So if you find the least square result unconvincing, squared distances are maybe not the adequate objective function for your problem. If you find an alternative one, gradient descent is a way to obtain at least a locally optimal solution.

Squared Euclidean distance [edit]

See also: Square (algebra)

The square of the standard Euclidean distance, which is known as the **squared Euclidean distance** (SED), is also of interest; as an equation:

$$d^2(\mathbf{p}, \mathbf{q}) = (p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_i - q_i)^2 + \cdots + (p_n - q_n)^2.$$

- $SSE = \sum (\hat{y}_i - y_i)^2 = (\hat{y}_1 - y_1)^2 + (\hat{y}_2 - y_2)^2 + (\hat{y}_3 - y_3)^2 + \cdots + (\hat{y}_n - y_n)^2$

=> SSE에 최소제곱법 적용 가능!!!