Suppose GPA can be predicted from the GMAT score. Consider the data for \( n \) students, where the GPA and GMAT of student \( i \in \{1, 2, \ldots, n\} \) are observed as \( y_i \) and \( x_i \). Use least squares method to estimate a hypothesized relation of the form

\[ y = a + bx. \]

\[ \Rightarrow \text{Error} = f(a, b) = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \]

Stationary Point \[ \nabla f = 0 \Rightarrow \begin{cases} \frac{\partial f}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0 \\ \frac{\partial f}{\partial b} = -2 \sum_{i=1}^{n} (y_i - a - bx_i)x_i = 0 \end{cases} \]

\[ \Rightarrow \begin{cases} \sum_{i} y_i = na + b \sum_{i} x_i \\ \sum_{i} x_i y_i = \sum_{i} x_i a + b \sum_{i} x_i^2 \end{cases} \Rightarrow \text{Two equations and two variables} \]

\( \Rightarrow \) Normally, there is only one solution.

(bad cases: no or infinitely many solutions)

- What is the type of stationary point?

\[ H(f(a, b)) = \begin{bmatrix} 2n & 2 \sum_{i} x_i \\ 2 \sum_{i} x_i & 2 \sum_{i} x_i^2 \end{bmatrix} \]
Study $H(f(a,b))$:

The leading principal minors of $H$ are: $2n > 0$ and

$$4n \geq \sum_{i} x_i^2 - 4(\sum x_i)^2 > 0$$

$g(x)$

It is well known that $g(x) > 0$ if at least two $x_i$ and $x_j$ are different.

Another approach: \( \min g(x) \rightarrow \nabla g(x) = 0 \)

\( \rightarrow g(x) > 0 \) and

$g(x) = 0$ if $x_1 = x_2 = ... = x_n$

\( (a,b) \) obtained from $\nabla f = 0$ are the best model relating GMAT to GPA if at least two entries of $x$ are different

\[
\min f(x) : x^* = \text{local} \Rightarrow f(x^*) > 0 \text{ and } f(x^*) > 0.
\]

We need to get some intuition about the proof to be able to design a numerical algorithm.

$$f(x^* + \Delta x) = f(x^*) + f'(x^*) \Delta x + \underbrace{\text{higher order terms}}_{O(\Delta x^2)}$$
This is Taylor series approximation.

Assume that \( f(x_*) \neq 0 \).

Then, there is a \( \Delta x \) such that \( f(x_*) \Delta x < 0 \).

Also, high-order terms are negligible compared to \( f(x_*) \Delta x \) if this first-order term is nonzero.

\[ \Rightarrow \text{If } f(x_*) \neq 0, \text{ then there exists a small perturbation } \Delta x \text{ such that } f(x_* + \Delta x) < f(x_*) \]

But since \( x_* \) is a local minimum, its perturbation should increase the function, i.e., \( f(x_* + \Delta x) > f(x_*) \) for small \( \Delta x \).

\[ \Rightarrow \text{This contradiction implies } f'(x_*) = 0. \]

How about \( f''(x_*) \)?

\[ f(x_* + \Delta x) = f(x_*) + f'(x_*) \Delta x + \frac{1}{2} f''(x_*) (\Delta x)^2 + \text{higher order terms} + O(\Delta x^3) \]

\[ \Rightarrow f(x_* + \Delta x) = f(x_*) + \frac{1}{2} f''(x_*) (\Delta x)^2 + \ldots \]
- If $f^{(1)}(x) < 0$, then $\frac{1}{2} f^{(2)}(x) \Delta x < 0$.

- Then, as before, $f(x_* + \Delta x) < f(x_*)$ for a small perturbation, which is contrary to $x_*$ being a local min.

$\Rightarrow f^{(2)}(x_*) < 0$

- Using Taylor series and above argument, we can say:

  - If $f^{(1)}(x_*) > 0$ and $f^{(1)}(x_*) = 0 \Rightarrow f(x_* + \Delta x) > f(x_*)$ for small $\Delta x$  

    $\Rightarrow x_*$ is local min

- If $f^{(1)}(x_*) = 0$ and $f^{(1)}(x_*) = \infty \Rightarrow$

  we should check higher derivative to check the status of $x_*$. 
\[
\min_{x \in \mathbb{R}^n} f(x) : x_0 = \text{local min} \implies \nabla f(x_0) = 0 \quad \text{and eigenvalues of} \quad H(f(x_0)) \geq 0.
\]

- We need some intuition.

- Taylor series in the multivariate case:

\[
f(x_0 + \Delta x) = f(x_0) + \nabla f(x_0) \Delta x + \ldots \]

- If \( \nabla f(x_0) \neq 0 \), then we can design a small vector \( \Delta x \) such that \( \nabla f(x_0) \Delta x < 0 \).

- This first-order term is more significant than the rest (residue)

- If \( \nabla f(x_0) \neq 0 \), then there is a small vector \( \Delta x \) such that \( f(x_0 + \Delta x) < f(x_0) \).

- But, \( x_0 \) is a local min and this can't happen.

\[
\implies \nabla f(x_0) = 0.
\]

- How about \( H(f(x_0)) \)?
\[ f(x_\ast + \Delta x) = f(x_\ast) + \nabla f(x_\ast) \Delta x + \frac{1}{2} (\Delta x)^T \mathbf{H}(f(x_\ast)) \Delta x \]

\[ + \ldots \]

\[ \text{high order terms} \]

\[ \Rightarrow f(x_\ast + \Delta x) = f(x_\ast) + \frac{1}{2} (\Delta x)^T \mathbf{H}(f(x_\ast)) \Delta x + \ldots \]

\[ \Rightarrow 1 - \frac{1}{2} (\Delta x)^T \mathbf{H}(f(x_\ast)) \Delta x \geq 0 \text{ for all small } \Delta x \]

if \( x_\ast \) is a local min.

2. If \( \frac{1}{2} (\Delta x)^T \mathbf{H}(f(x_\ast)) \Delta x \geq 0 \) for all small \( \Delta x \)

and \( \nabla f(x_\ast) = 0 \) \( \Rightarrow x_\ast \) = local min

3. If \( \frac{1}{2} (\Delta x)^T \mathbf{H}(f(x_\ast)) \Delta x = 0 \) for some small \( \Delta x \)

we should work on high-order terms.

4. If \( \frac{1}{2} (\Delta x)^T \mathbf{H}(f(x_\ast)) \Delta x < 0 \) for some small \( \Delta x \),

then \( x_\ast \) can't be a local min.

- How to check sign of \( \Delta x^T \mathbf{H}(f(x_\ast)) \Delta x \) for all small \( \Delta x \)?