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

$$\Rightarrow \quad \text{error} = f(a,b) = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

$$\xrightarrow[\text{point}]{\text{stationary}} \quad \nabla f = 0 \quad \Rightarrow \quad \begin{cases} \dfrac{\partial f}{\partial a} = -2 \displaystyle\sum_{i=1}^{n} (y_i - a - bx_i) = 0 \\ \dfrac{\partial f}{\partial b} = -2 \displaystyle\sum_{i=1}^{n} (y_i - a - bx_i) x_i = 0 \end{cases}$$

$$\Rightarrow \quad \begin{cases} \sum_i y_i = na + b \sum_i x_i \\ \sum_i x_i y_i = \sum_i x_i a + b \sum_i x_i^2 \end{cases} \quad \Rightarrow \quad \text{Two equations and two variables}$$

$\Rightarrow$ 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 $H(f(a,b))$:

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

$$\underbrace{4n \sum_i x_i^2 - 4\left(\sum x_i\right)^2}_{g(x)} \overset{?}{>} 0$$

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

- Another approach: $\min_x g(x) \longrightarrow \nabla g(x) = 0$

$\longrightarrow g(x) \geq 0$ and

$g(x) = 0$ if $x_1 = x_2 = \cdots = x_n$

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

---

$\min_{x \in \mathbb{R}} f(x)$: $x_* =$ local $\Rightarrow$ $f'(x_*) = 0$ and $f''(x_*) \geq 0$

we need to get some intituition about the proof to be able to design a numerical algorithm.

$$f(x_* + \Delta x) = f(x_*) + f'(x_*)\,\Delta x + \underbrace{\text{higher order terms}}_{O(\Delta x^2)}$$

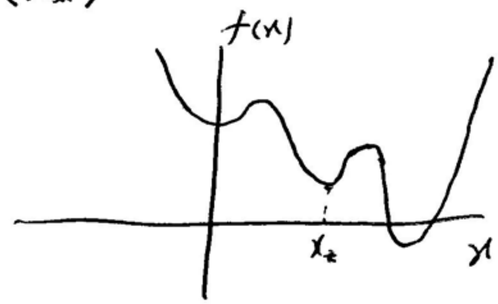
- This is Taylor series approximation.

- Assume that $f'(x_*) \neq 0$.

- Then, there is a Δx 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

$\Rightarrow$ If $f'(x_*) \neq 0$, then there exists a small perturbation Δx 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 Δx

$\Rightarrow$ This contradiction implies $f'(x_*) = 0$

- How about $f''(x_*)$?

$$f(x_* + \Delta x) = f(x_*) + \cancel{f'(x_*)}^{\,0}\Delta x + \frac{1}{2} f''(x)(\Delta x)^2 + \underbrace{\text{higher order terms}}_{O(\Delta x^3)}$$

$$\Rightarrow f(x_* + \Delta x) = f(x_*) + \frac{1}{2} f''(x)(\Delta x)^2 + \cdots$$

- If $f''(x) < 0$, then $\frac{1}{2} f''(x) \Delta x < 0$.
- Then, as before, $f(x_* + \Delta x) < f(x_*)$ for a small perturbation, which is contarary to $x_* =$ Local min

$\Rightarrow f''(x_*) \geq 0$

- Using Taylor series and above argument, we can say:
  - If $f''(x_*) > 0$ and $f'(x_*) = 0 \Rightarrow$
    $f(x_* + \Delta x) > f(x_*)$ for small Δx
    $\Rightarrow x_* =$ Local min
  - If $f''(x_*) = 0$ and $f'(x_*) = 0 \Rightarrow$
    we should check higher derivative to check the status of x_*.

$\min_{x \in \mathbb{R}^n} f(x)$: x_* = local min

$$\Rightarrow \quad \nabla f(x_*) = 0 \quad \text{and eigenvalues of } H(f(x_*)) \geq 0$$

- We need some intuition.

- Taylor series in the multivariate case:

$$f(\underbrace{x_* + \Delta x}_{\text{perturbation} \in \mathbb{R}^n}) = f(x_*) + \underbrace{\nabla f(x_*)}_{\text{row vector}} \underbrace{\Delta x}_{\text{column vector}} + \underbrace{\cdots\cdots}_{\text{high order terms}}$$

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

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

- If $\nabla f(x_*) \neq 0$, then there is a small vector Δx such that $f(x_* + \Delta x) < f(x_*)$

- But, x_* is a local min and this can't happen.

$$\Rightarrow \quad \nabla f(x_*) = 0$$

- How about $H(f(x_*))$?

$$f(x_* + \Delta x) = f(x_*) + \cancel{\nabla f(x_*)}^{\,0} \Delta x + \frac{1}{2} \underbrace{(\Delta x)^T}_{\text{row}} \underbrace{H(f(x_*))}_{\text{matrix}} \underbrace{\Delta x}_{\text{column}}$$

$$+ \underbrace{\cdots\cdots}_{\text{high order terms}}$$

$$\Rightarrow \quad f(x_* + \Delta x) = f(x_*) + \frac{1}{2} (\Delta x)^T H(f(x_*)) \Delta x + \cdots$$

$\Rightarrow$ 1- $\frac{1}{2} (\Delta x)^T H(f(x_*)) \Delta x \geq 0$ for all small Δx if x_* is a local min.

2- If $\frac{1}{2} (\Delta x)^T H(f(x_*)) \Delta x > 0$ for all small Δx and $\nabla f(x_*) = 0$ $\Rightarrow$ x_* = local min

3- If $\frac{1}{2} (\Delta x)^T H(f(x_*)) \Delta x = 0$ for some small Δx (nonzero) we should work on high-order terms.

4- If $\frac{1}{2} (\Delta x)^T H(f(x_*)) \Delta x < 0$ for some small Δx, then x_* can't be a local min.

- How to check sign of $\Delta x^T H(f(x_*)) \Delta x$ for all small Δx ?