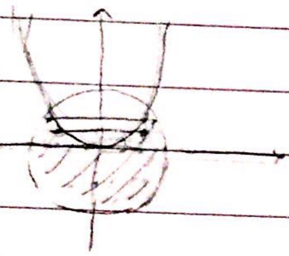
سوال ۱:

آ) می‌دانیم اگر مشتق دوم مثبت باشد تابع convex است ← $5x^4 + 3x^3 \to 20x^3 + 9x^2 \to$

$60x^2 + 18x = 6(10x^2 + 3x) \to 60(x^2 + \frac{3}{10}x)$

$\Rightarrow 60(x^2 + \frac{3}{10}x + \frac{9}{400} - \frac{9}{400}) = 60(x + \frac{3}{20})^2 - \frac{9\times6}{40} = 60(x + \frac{3}{20})^2 - \frac{27}{20}$

به ازای برخی مقادیر x منفی است ← تابع محدب نیست



ب) با توجه به شکل تابع می‌دانیم اگر دو نقطه‌ی نشان داده شده در شکل را ببینیم تابع محدب نخواهد بود

$x^2 + y^2 \le 1 \to y^2 \le 1 - x^2 \quad \text{if } y \le x^2$

ج) طبق تعریف می‌دانیم تابع convex است اگر $f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y)$:

$\sum_i^n |x_i| \ge 0 \quad , \quad \sum_i^n \theta|x_i| = \theta \sum_i^n |x_i|$

با توجه به نامساوی بالا اگر x_1 , x_2 دو عضو از مجموعه باشند طبق نامساوی مثلث‌ها داریم:

$\sum_i^n \theta|x_{1i}| + (1-\theta)|x_{2i}| \ge \sum_i^n |\theta x_{1i} + (1-\theta)x_{2i}|$ ← که همان تعریف convex است.

د) برای اینکه ثابت کنیم convex است باید ثابت کنیم مشتق دوم آن همواره مثبت است به شرطی که A لازم باشد

$$x^T A x = [x_1 \dots x_n] \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [(x_1A_{11} + x_2A_{21} + \dots + x_nA_{n1}) \dots$$

$$(x_1A_{1n} + x_2A_{2n} + \dots + x_nA_{nn})] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1^2 A_{11} + x_1x_2A_{21} + \dots + x_1x_nA_{n1} + \dots +$$

$$x_nx_1A_{1n} + x_nx_2A_{2n} + \dots + x_n^2A_{nn})$$

که اگر ماتریس هسین این را حساب کنیم چون می‌دانیم A_{ij} ها مثبت اند مثبت می‌شود.

s.a.m

سوال ۲:

۲)

$$\frac{\partial f(w_0, w_1, \dots, w_n)}{\partial w_0} + \frac{\partial f(w_0, \dots, w_n)}{\partial w_1} + \dots + \frac{\partial f(w_0, \dots, w_n)}{\partial w_n}$$

$$\Rightarrow \frac{\partial [\sum_i w_n x_i^n + w_{n-1} x_i^{n-1} + \dots + w_0 x_i^0 - y^{(i)}]^2}{\partial w_0} = 2 \sum_i w_n x_i^n + \dots + w_0 - y^i$$

$$+ \frac{\partial [\sum_i w_n x_i^n + w_{n-1} x_i^{n-1} + \dots + w_0 x_i - y^i]^2}{\partial w_1} = 2x \sum_i w_n x_i^n + \dots + w_0 - y^i$$

$$\vdots$$

$$+ \frac{\partial [\sum_i w_n x_i^n + \dots + w_0 x^0 - y^i]^2}{\partial w_n} = 2x^n \sum_i w_n x_i^n + \dots + w_0 - y^i$$

$$= 2 [1, x_i, \dots, x_i^n] \sum_i \hat{y}^i - y^i$$

ب) گرادیان بر حسب w_i ها را درست مثل حساب کردیم حالا $w_i^{(0)}$ هایی برای شروع انتخاب می‌کنیم و آن را در رابطه ∇f حساب می‌کنیم

$$w - \alpha \nabla f \Rightarrow \begin{bmatrix} w_0^0 - 2\alpha \sum_i \hat{y}^i - y^i \\ w_1^0 - 2x\alpha \sum_i \hat{y}^i - y^i \\ \vdots \\ w_n^0 - 2x^n \alpha \sum_i \hat{y}^i - y^i \end{bmatrix} = \begin{bmatrix} w_0^1 \\ \vdots \\ w_n^1 \end{bmatrix}$$

این فرایند را آنقدر ادامه می‌دهیم تا تابع خطا مینیمم شود.

حالا α هرچقدر بزرگ‌تر باشد کاهش در هر مرحله بزرگ‌تر می‌شود و اگر خیلی بزرگ باشد ممکن است از نقطه مینیمم رد شود
و اگر α را کاهش دهیم سرعت همگرا شدن تابعمان کم می‌شود و با تکرارهای بیشتر به جواب نهایی می‌رسیم

s.a.m