**Problem 1.**
Using the definition of a convex function, let $\gamma = c\theta_1 + (1-c)\theta_2 \ \forall c \in [0,1]$. Then we can set

$$\begin{aligned} f(\theta_1) &\geq f(\gamma) + \nabla f(\gamma)^T(\theta_1 - \gamma) \text{ and} \\ f(\theta_2) &\geq f(\gamma) + \nabla f(\gamma)^T(\theta_2 - \gamma) \end{aligned}$$

We can multiply the inequality involving $\theta_1$ by $c$ and the one involving $\theta_2$ by $(1-c)$.

$$\begin{aligned} cf(\theta_1) &\geq cf(\gamma) + c\nabla f(\gamma)^T(\theta_1 - \gamma) \\ (1-c)f(\theta_2) &\geq (1-c)f(\gamma) + (1-c)\nabla f(\gamma)^T(\theta_2 - \gamma) \end{aligned}$$

We can add the two lines to combine into one inequality.

$$\begin{aligned} cf(\theta_1) + (1-c)f(\theta_2) &\geq cf(\gamma) + c\nabla f(\gamma)^T(\theta_1 - \gamma) + (1-c)f(\gamma) + (1-c)\nabla f(\gamma)^T(\theta_2 - \gamma) \\ cf(\theta_1) + (1-c)f(\theta_2) &\geq cf(\gamma) + f(\gamma) - cf(\gamma) + \nabla f(\gamma)^T\left[c(\theta_1 - \gamma) + (1-c)(\theta_2 - \gamma)\right] \\ cf(\theta_1) + (1-c)f(\theta_2) &\geq f(\gamma) + \nabla f(\gamma)^T\left[c\theta_1 - c\gamma + \theta_2 - \gamma - c\theta_2 + c\gamma\right] \\ cf(\theta_1) + (1-c)f(\theta_2) &\geq f(\gamma) + \nabla f(\gamma)^T\left[c\theta_1 + (1-c)\theta_2 - \gamma\right] \\ cf(\theta_1) + (1-c)f(\theta_2) &\geq f(\gamma) + \nabla f(\gamma)^T\left[c\theta_1 + (1-c)\theta_2 - [c\theta_1 + (1-c)\theta_2]\right] \\ cf(\theta_1) + (1-c)f(\theta_2) &\geq f(\gamma) \\ cf(\theta_1) + (1-c)f(\theta_2) &\geq f(c\theta_1 + (1-c)\theta_2) \ \blacksquare \end{aligned}$$

Which is convex by definition.
**Problem 2.**
**a.**
Our $R^2$ for our training and test models are 0.510 and 0.505 respectively.
**b.**
The predicted price for the house is $15,436,769.54. Given that this house is owned by one of the richest men in the world and is on a huge lot, this price is reasonable.
**c.**
Our $R^2$ for our training and test models with feature engineering are 0.517 and 0.511 respectively.
**d.**
Using gradient descent, our $R^2$ for our training and test models are 0.510 and 0.505 respectively. Our $R^2$ for our training and test models with feature engineering are 0.517 and 0.511 respectively. This does not come as surprising as we expect our gradient descent to give very close approximations to the $\beta$ vector as our closed-form OLS solution.
The predicted model for Bill Gates House is $15,436,770.00, also extremely close to our original prediction
**e**
Using gradient descent, our $R^2$ for our training and test models are 0.510 and 0.505 respectively. Our $R^2$ for our training and test models with feature engineering are 0.501 and 0.506 respectively. These accuracies are slightly less than our regular gradient descent, perhaps because our randomness of the stochastic gradient descent did not capture a good enough sample through each iteration.
The predicted model for Bill Gates House is $15,546,210.00, also extremely close to our original prediction, just slightly larger.
**Problem 3.**
A function $f$ is $\mu$-strongly convex if $g(x) = f(x) - \frac{\mu}{2}||x||_2^2$ is convex. The function $g(x)$ being convex implies $\nabla g(x)$ is monotone. Formally:

$$\begin{aligned} (\nabla g(\theta_1) - \nabla g(\theta_2))^T(\theta_1 - \theta_2) &\geq 0 \\ (\nabla f(\theta_1) - \mu\theta_1 - \nabla f(\theta_2) - \mu\theta_2)^T(\theta_1 - \theta_2) &\geq 0 \\ \langle \nabla f(\theta_1) - \nabla f(\theta_2) - \mu(\theta_1 - \theta_2), \theta_1 - \theta_2 \rangle &\geq 0 \\ \langle \nabla f(\theta_1) - \nabla f(\theta_2), \theta_1 - \theta_2 \rangle &\geq \mu||\theta_1 - \theta_2||_2^2 \ \blacksquare \end{aligned}$$

We now aim to solve the recursion, that is (Assuming $M_g = 0$):

$$\mathbb{E}\big[||\theta^{(t+1)} - \theta^*||_2^2\big] \leq (1 - 2\mu\eta_t + \eta_t^2 M_g L^2)\mathbb{E}\big[||\theta^{(t)} - \theta^*||_2^2\big] + \eta_t^2\sigma_g^2$$

$$\leq (1 - 2\mu\eta_t)\mathbb{E}\big[||\theta^{(t)} - \theta^*||_2^2\big] + \eta_t^2\sigma_g^2$$

$$\leq (1 - 2\mu\eta_t)\mathbb{E}\big[||\theta^{(t)} - \theta^*||_2^2\big] + \eta_t^2\sigma_g^2 \leq (1 - 2\mu\eta_t)\big[(1 - 2\mu\eta_{t-1})\mathbb{E}\big[||\theta^{(t-1)} - \theta^*||_2^2\big] + \eta_t^2\sigma_g^2\big] + \eta_{t-1}^2\sigma_g^2$$

$$(1 - 2\mu\eta_t)\mathbb{E}\big[||\theta^{(t)} - \theta^*||_2^2\big] + \eta_t^2\sigma_g^2 \leq (1 - 2\mu\eta_t)(1 - 2\mu\eta_{t-1})\dots(1 - 2\mu\eta_1)\mathbb{E}\big[||\theta^{(0)} - \theta^*||_2^2\big] + \sum_{i=1}^{t}\eta_i^2\sigma_g^2$$

$$(1 - 2\mu\eta_t)\mathbb{E}\big[||\theta^{(t)} - \theta^*||_2^2\big] \leq (1 - 2\mu\eta_t)(1 - 2\mu\eta_{t-1})\dots(1 - 2\mu\eta_1)\mathbb{E}\big[||\theta^{(0)} - \theta^*||_2^2\big] - \eta_t^2\sigma_g^2 + \sum_{i=1}^{t}\eta_i^2\sigma_g^2$$

$$\mathbb{E}\big[||\theta^{(t)} - \theta^*||_2^2\big] \leq (1 - 2\mu\eta_{t-1})\dots(1 - 2\mu\eta_1)\mathbb{E}\big[||\theta^{(0)} - \theta^*||_2^2\big] + \frac{\sum_{i=1}^{t-1}\eta_i^2\sigma_g^2}{(1 - 2\mu\eta_t)}$$

$$\mathbb{E}\big[||\theta^{(t)} - \theta^*||_2^2\big] \leq (1 - 2\mu\eta_{t-1})\dots(1 - 2\mu\eta_1)\mathbb{E}\big[||\theta^{(0)} - \theta^*||_2^2\big] + \frac{\sum_{i=1}^{t-1}\eta_i^2\sigma_g^2}{(1 - 2\mu\frac{c}{t+1})}$$

Which is now of the form: $\mathbb{E}\big[||\theta^{(t)} - \theta^*||_2^2\big] \leq \dfrac{c_0}{1+t}$

**Problem 4.**
We have $f(\theta) = ||\theta||_2$. The sub-gradient of this function is

$$\partial f(\theta) = \begin{cases} \frac{1}{||\theta||_2}\theta, \ \theta = 0 \\ \text{any vector } \gamma \text{ in } \mathbb{R}^d, ||\gamma||_2 \leq 1, \ \theta = 0 \end{cases}$$

**Pledge:**
Please sign below (print full name) after checking (✓) the following. If you can not honestly check each of these responses, please email me at kbala@ucdavis.edu to explain your situation.

- I pledge that I am an honest student with academic integrity and I have not cheated on this homework.
- These answers are my own work.
- I understand that to submit work that is not my own and pretend that it is mine is a violation of the UC Davis code of conduct and will be reported to Student Judicial Affairs.
- I understand that suspected misconduct on this homework will be reported to the Office of Student Support and Judicial Affairs and, if established, will result in disciplinary sanctions up through Dismissal from the University and a grade penalty up to a grade of "F" for the course.

Name: Gianni Spiga