General Note: Most solutions are easy if results of the previous questions are used.

1. Let S be a non-empty subset of $\mathbf{R}^n$. S is said to be convex if for any $x, y \in S$, $\alpha x + \beta y \in S$ holds true for any choice of $\alpha, \beta \in [0,1]$ such that $\alpha + \beta = 1$. Let $x, y \in S$ and let $z = \alpha x + \beta y$ satisfy the conditions of the definition above. Further suppose that $||x - z|| = s > 0$ and $||y - z| = t > 0$. Show that $\alpha = \frac{t}{s+t}$ and $\beta = \frac{s}{s+t}$.

2. Let A be an $m \times n$ real matrix. Let b be a vector in $\mathbf{R}^m$. Let $S = \{x \in \mathbf{R}^n : Ax \leq b\}$ and let $T = \{y \in \mathbf{R^m} : y = Ax \text{ for some } x \in \mathbf{R}^n\}$. Are S and T convex sets?

3. Let S and T be non-empty convex subsets of $\mathbf{R}^n$.
    1. Give a counterexample to show that $S \cup T$ need not be convex.
    2. Show that $S \cap T$ is convex.
    3. Is the set $S + T = \{z : z = x + y \text{ for some } x \in S, y \in T\}$ convex? Either prove the statement or give a counter example.

4. Let S be a convex subset of $\mathbf{R}^n$. Let $f : S \mapsto \mathbf{R}$ be a function. f is convex if for any $x, y \in S$ and $\alpha, \beta \in [0,1]$ such that $\alpha + \beta = 1$, $f(\alpha x + \beta y) \leq \alpha f(x) + \beta f(y)$. Show that if $\alpha, \beta\ \gamma$ are positive real numbers such that $\alpha + \beta + \gamma = 1$ and if $x, y, z \in S$, then $f(\alpha x + \beta y + \gamma z) \leq \alpha f(x) + \beta f(y) + \gamma f(z)$.

5. Let S be a convex subset of $\mathbf{R}^n$. Let $f : S \mapsto \mathbf{R}$ be a function. f is *concave* if for any $x, y \in S$ and $\alpha, \beta \in [0,1]$ such that $\alpha + \beta = 1$, $f(\alpha x + \beta y) \geq \alpha f(x) + \beta f(y)$. Show that f is concave if and only if $-f$ is convex. Let a be any vector in $\mathbf{R}^n$ and b a real number. Show that the function defined by $f(x) = a^T x + b$ is both convex and concave.

6. Let S be a convex subset of $\mathbf{R}^n$. Let $f : S \mapsto \mathbf{R}$ be a convex function. Show that $f(S)$ defined by: $f(S) = \{y : y = f(x) \text{ for some } x \in S\}$ a convex set always? Either prove the statement or give a counter example.

7. Let $f : \mathbf{R} \mapsto \mathbf{R}$ be convex. Let $x < z < y$ be three distinct points on the real line.
    1. Show that $\frac{f(z)-f(x)}{z-x} \leq \frac{f(y)-f(x)}{y-x}$. Prove the converse as well - that is, if f satisfies this condition for all $x < z < y$, then f must be convex. (Hint: Start with defining d by $d = y - x$. Now let $\epsilon > 0$ be chosen so that $z - x = \epsilon d$. The rest is easy.)
    2. Show that $\frac{f(z)-f(x)}{z-x} \leq \frac{f(y)-f(z)}{y-z}$. Prove the converse as well - that is, if f satisfies this condition for all $x < z < y$, then f must be convex. (Hint: Similar to the first question.)
    3. Suppose further that f is differentiable everywhere. Then, show that $f'(x) \leq f'(y)$. (That is, the derivative is monotonically increasing.) (Hint: In the first sub question, note that as $\alpha \to 1, z \to x+$ and as $\alpha \to 0, z \to y-$ and so on.)
    4. To Prove the converse - that is, if f is differentiable everywhere and if f' is an increasing function on x, then f is convex. (Hint: This requires Mean Value Theorem and subquestion (2).)
    5. Show that if f is convex and differentiable, then $f(y) \geq f(x) + f'(x)(y - x)$ for every pair of points x, y in $\mathbf{R}$.
    6. Show that if for all $x, y \in \mathbf{R}$, $f(y) \geq f(x) + f'(x)(y - x)$, then f is convex. (Hint: Let $x, y \in \mathbf{R}$ and let $z = \alpha x + \beta y$ for some non negative α, β satisfying $\alpha + \beta = 1$. By assumption, we have $f(x) \geq f(z) + f'(z)(x - z)$ and $f(y) \geq f(z) + f'(z)(y - z)$. Now consider $\alpha f(x) + \beta f(y)$ and simplify the RHS).
    7. Suppose f is twice differentiable and convex. Then show that f'' is non-negative everywhere. (Hint: Use the result of subquestion (4).)
    8. Prove the converse. That is, if f is twice differentiable and $f''(x) \geq 0$ everywhere. Then f is convex. (Hint: This requires analysizing the second order term in the Taylor expansion of f around x.)

8. Determine whether the following functions over $\mathbf{R}$ are convex, concave or neither. (Whenever the function is differentiable, then it is simpler to use that fact. You will see that the previous exercise gave immensely powerful tools!)
(1). $f(x) = \log x$, (2). $f(x) = p|x| + q$ for some constants p and q (you will see methods to solve this question easily later. But try it now!) (3). $f(x) = x^2$, (4). $f(x) = \frac{1}{1+e^{-x}}$, (5). $f(x) = \log(1 + e^{-x})$.

9. Show that the norm function $f : \mathbf{R}^n \mapsto \mathbf{R}$ defined by $f(x) = ||x||^2$ is convex. (Hint: It is enough to show that for any positive α, β satisfying $\alpha + \beta = 1$, $\alpha||x||^2 + \beta||y||^2 - ||\alpha x + \beta y||^2$ is non-negative. Now use the fact that $||x+y||^2 = ||x||^2 + ||y||^2 + 2\,(x,y)$, where (x,y) is the inner product of x and y).

10. Let A be an $m \times n$ matrix and $b \in \mathbf{R}^m$. Let $f : \mathbf{R^m} \mapsto \mathbf{R}$ be a convex function. Then, show that $g(x) = f(Ax+b)$ is a convex function from $\mathbf{R^n}$ to $\mathbf{R}$. This exercise shows that composition of an affine function with a convex function is convex. An important special case is when $m = 1$. Then, $g(x) = f(a^T x + b)$ for some $a \in \mathbf{R}^n$ and $b \in \mathbf{R}$, and a convex function $f : \mathbf{R} \mapsto \mathbf{R}$.

11. Recall that a function $f : \mathbf{R} \mapsto \mathbf{R}$ is monotone (increasing) if for all $x, y \in \mathbf{R}$, if $x < y$ then $f(x) \leq f(y)$. Let $f : \mathbf{R} \mapsto \mathbf{R}$ be both monotone and convex. Let $g : \mathbf{R}^n \mapsto \mathbf{R}$ be convex. Then, show that the function $h(x) = f(g(x))$ is convex. (Where did you use the monotonicity of f in the proof?)

12. Given an example for convex functions $f, g : \mathbf{R} \mapsto \mathbf{R}$ such that $h(x) = f(g(x))$ is not convex. (Note: The point of this exercise is to show that unless monotonicity of f is assumed, the result of the previous question will not hold.)

13. Let $f, g : \mathbf{R}^n$ to $\mathbf{R}$ be convex. Let s, t be any positive real numbers. Show that $h(x) = sf(x) + tg(x)$ is convex. (This allows us to combine convex functions to form new ones.)

14. Now we are ready to use the apparatus developed in the preceding questions to prove the convexity of some important functions in machine learning.

    1. *Squared Loss Function:* Let A be an $m \times n$ matrix and $b \in \mathbf{R}^m$. Define $f : \mathbf{R}^n \mapsto \mathbf{R}$ by $f(\theta) = ||b - A\theta||^2$. Show that f is convex. (Hint: Don't try to solve the problem from scratch. The results proved in previous questions will be sufficient to readily solve the question!)
    2. *Regularized Regression function:* Let A, b be as defined in the previous subquestion. Let $\lambda > 0$. Show that $f(\theta) = ||b - A\theta||^2 + \lambda||\theta||^2$ is convex. (Hint: Same comment as the previous question.)
    3. *Logistic Regression Function:* Let $a \in \mathbf{R}^n$ and $b \in \mathbf{R}$. Show that the function $f : \mathbf{R}^n \mapsto \mathbf{R}$ defined by $f(\theta) = \log\left(1 + e^{a^T\theta + b}\right)$ is convex. (Hint: Of course, don't start from scratch!)

15. The Rolle's Theorem states that if $f : \mathbf{R} \mapsto \mathbf{R}$ is differentiable in the open interval $[a, b]$ for some $a < b$ such that $f(a) = f(b) = 0$ for $a, b \in \mathbf{R}, a \neq b$, then there exists $c \in (a, b)$ such that $f'(c) = 0$. Suppose that f is twice differentiable in $[a, b]$. Further suppose that $f(a) = f'(a) = f(b) = 0$. Show that there exists points d, c satisfying $a < d < c < b$ such that $f'(c) = 0$ and $f''(d) = 0$.

16. Assume that a function $f : \mathbf{R} \mapsto \mathbf{R}$ be n times differentiable. Let $a \in \mathbf{R}$. Show that if f can be written in the form $f(x) = a_0 + a_1(x-a) + a_2(x-a)^2 + \cdots a_n f(x-a)^n$, then we must have $a_0 = f(a), a_1 = f'(a), a_2 = \frac{1}{2!}f''(a) \cdots, a_n = \frac{1}{n!}f^{(n)}(a)$, where $f^{(k)}(a)$ denotes the k^{th} derivative of f at a. (This exercise shows that if a function has f has a Taylor expansion, it can happen only one way! If f is a polynomial of degree n, the above set of n equations should be sufficient to determine $a_0, a_2 \ldots a_n$. When, f is not a polynomial, the expansion can provide only an approximation.)

17. Rolle's theorem allows us to develop the following form of Taylor approximation. The issue with this form is that it involves the second derivative of f at an unknown point d near a, and hence gives a mathematical charecterization that hard from a computational point of view. However, the theorem helps us to derive error bouns for many practical algorithms. Let a, b be points on the real line with $a < b$. Let $f : \mathbf{R} \mapsto \mathbf{R}$.

    1. Suppose that f is differentiable in (a, b). Then use Rolle's theorem to show that there exists c satisfying $a < c < b$ such that $f(b) = f(a) + f'(a)(b-a)$. (Hint: Find coefficients a_0, a_1 such that the function $g(x)$ satisfying $g(x) = f(x) - a_0 - a_1(x-a)$, satisfies the Rolle's condition $g(a) = g(b) = 0$.)
    2. Assuming that $f(x)$ is a differentiable twice, show that there exists a point $d \in (a, b)$ such that $f(b) = f(a) + f'(a)(b-a) + \frac{f''(d)}{2}(b-a)^2$.

18. This questions extends the results of the previous question to multivariable functions. Let $f : \mathbf{R}^n \mapsto \mathbf{R}$ be be differentiable twice everywhere in the domain. Let $a, b \in \mathbf{R}^n$, $a \neq b$. Define the function $g : [0, 1] \mapsto \mathbf{R}$ by: $g(t) = f(a + t(b-a))$. For $0 \leq t \leq 1$ Define $G(t) = \nabla f(a + t(b-a))$ and $H(t)$ to be the Hessian of f at the point $a + t(b-a)$. (Note that $H(t)$ is not defined as the Hessian of G at t.) Note further that $G(t)$ is a row vector and $H(t)$ is an $n \times n$ matrix.

1. Let $t \in (0,1)$. Find the derivative and second derivative of G using the chain rule of differentiation.
2. Now apply the previous question's results to show that there exists points c, d and $\alpha, \beta \in (0,1)$ such that $f(b) = f(a) + G(\alpha)(b-a) = f(a) + G(0)(b-a) + \frac{1}{2}(b-a)^T H(\beta)(b-a)$.

19. Let $f : \mathbf{R}^n \mapsto \mathbf{R}$ be any function. Let G_f be the subset of $\mathbf{R}^{n+1}$ defined by:
$G_f = \{(x,y) : x \in \mathbf{R}^n, y \in \mathbf{R} \text{ and } f(x) \leq y\}$. Show that the set G_f is a convex subset of $\mathbf{R}^{n+1}$ if and only if f is a convex function over $\mathbf{R}^n$.

20. Show that if $f, g : \mathbf{R}^n \mapsto \mathbf{R}$ are convex functions, then $h(x) = \max\{f(x), g(x)\}$ is convex. (Hint: Use the criteria proved in the previous question.).

21. Show that the function $g : \mathbf{R} \mapsto \mathbf{R}$ given by $g(x) = |x|$ is convex. (Hint: $|x|$ is the larger number among x and $-x$. Now use what you proved earlier.)

22. Prove that if $a \in \mathbf{R}^n$ and $b \in \mathbf{R}$, the function $f : \mathbf{R}^n \mapsto \mathbf{R}$ defined by $f(x) = |a^T x + b|$ is convex. (Hint: Of course use previous results, never start from scratch!)

23. Let S be a convex subset of $\mathbf{R}^n$. Let $f : S \mapsto \mathbf{R}$ be a convex function. Let $\epsilon > 0$ be a positive real number. Suppose that $x_0 \in S$ be a point such that $f(z) \geq f(x_0)$ for all $z \in S$ satisfying $||z - x_0|| < \epsilon$, then prove that for all $y \in S$, $f(x_0) \leq f(y)$. That is, if x_0 is a local minimum of f in S, then x_0 is a global minimum of f. Where did you use the fact that S is convex in proving the result?

24. Give an example of a convex function $f : \mathbf{R} \mapsto \mathbf{R}$ and a non empty subset S of $\mathbf{R}$ such that a local minimum of f in S is not a global minimum of f in S. (Note: The point here is that the previous result does not hold when S is not convex.)

25. If $f : \mathbf{R}^n \mapsto \mathbf{R}$ is convex and differentiable, then show that for all $x, y \in \mathbf{R}^n$, $f(y) \geq \nabla f(x)(y-x)$. (Hint: Start with the definition of convex functions that $f(y + \epsilon(y-x)) \leq f(x) + \epsilon(f(y) - f(x))$ for any $\epsilon > 0$. Now observe that as $\epsilon \to 0$, the directional derivative $\frac{f(y+\epsilon(y-x))-f(x)}{\epsilon} \to (\nabla f(x), y-x)$.

26. Prove the converse for the previous question. That is, if f is differentiable everywhere in $\mathbf{R}^n$ and suppose for all $x, y \in \mathbf{R}^n$ the condition $f(y) \geq f(x) + \nabla f(x)(y-x)$ holds, then show that f is convex. (Hint: You have seen the proof strategy in an earlier question!)

27. Suppose that $f : \mathbf{R}^n \mapsto \mathbf{R}$ is convex and differentiable, then x_0 is a global minimum of f **if and only if** $\nabla f(x_0) = 0$. (Hint: Use the result in the previous question.)

28. Suppose $f : \mathbf{R}^n \mapsto \mathbf{R}$ is twice differentiable with the second order partial derivatives being continuous. For each $x \in \mathbf{R}^n$, let $H_f(x)$ be the Hessian matrix of f at x.

    1. Show that if $H_f(x)$ is positive definite for all x, then f is convex. (Hint: Use the second order Taylor expansion of f.)
    2. For proving the converse, show that if there exists some $x, d \in \mathbf{R}^n$ such that $d^T H_f(x) d < 0$, then f is not convex. (Hint: Use the fact that for points $x + \epsilon d$ with small positve ϵ, due to the continuity of H_f, $d^T H_f(x + \epsilon d) d < 0$)).
    3. Suppose there exists a point $x_0 \in \mathbf{R}^n$ such that the Hessian H_f has an negative Eigen value λ with a corresponding non-zero Eigen vector d at, then f will not be convex. (Hint: Here you have to use the fact that the determinent function ($\mathbf{R}^{n \times n} \mapsto \mathbf{R}$) is a continuous function on the coefficients of the argument matrix. Hence, if the second order partial derivatives of f are continuous at x_0, then moving a small distance from x_0 will not change the componenets of the Hessian matrix by a large value, and due to the continuity of the determinent, the change in the Eigen values will also be very small. Consequently, $H_f(x)$ will have some negative Eigen value at each points x in a small neighborood around x_0, although the Eigen values and the corresponding Eigen vectors may slightly differ. As a final point, the continuity of the determinent in its input arguments is easy to see, as the determinent function involves only addition and multiplication operations on the coefficients of the matrix.)

29. Let A be an $m \times n$ matrix and $b \in \mathbf{R}^m$. Define $f : \mathbf{R}^n \mapsto \mathbf{R}$ by $f(\theta) = ||b - A\theta||^2$. Show that if for all $x \in \mathbf{R}^n$, $Ax \neq 0$ except when $x = 0$, then for every $d \in \mathbf{R}^n$ $d^T H_f(\theta) d > 0$. That is, the Hessian is positive definite.

30. Consider the Gradient descent algorithm applied to a convex function over a convex set, executing T iterations. Suppose that, instead of the update rule $x_{t+1} = x_t - \gamma \nabla f(x_t)$ we use a new rule: $x_{t+1} = x_t - \gamma y_t$, where $y_0, y_1, \ldots y_{T-1}$ are some arbitrary real vectors. Let x^* be the optimal solution. show that $\sum_{t=0}^{T-1} (x_t - x^*, y_t) \leq \frac{||x^*||^2}{2\gamma} + \frac{\gamma}{2} \sum_{i=0}^{T-1} ||y_t||^2$.

31. Given $x_1, x_2, \ldots x_m \in \mathbb{R}^n$ and labels $y_1, y_2 \ldots y_m \in \{\pm 1\}^n$, the perceptron algorithm works as the following:

    - Initialize: $t = 0; \theta_t = 0^m$;
    - While (1)
        - if there is some x_i such that $y_i(\theta_t, x_i) < 0$
        - then $\theta_{t+1} = \theta_t + y_t x_i$;
        - else return θ_t;
        - $t = t + 1$;
    - endWhile;

    We showed that If there exists some $\theta \in \mathbf{R}^m$ such that $y_i(\theta, x_i) > 0$ for all i, then this algorithm will converge to one such vector θ in at most $(RB)^2$ iterations, where $B = \min\{||\theta|| : y_i(\theta, x_i) > 0 \text{ for all } i\}$ and $R = \max_i ||x_i||$. Suppose that we change the update rule as $\theta_{t+1} = \theta_t + \gamma y_i x_i$ for some constant $\gamma > 0$, will there be any change in the number of iterations required for the algorithm to terminate?