Q1: $\inf\{sin(n) : n \geq 1\} = ?$ write down the answer and prove it

$\inf\{\sin(n) : n \geq 1\} = -1$，证明:

首先，$sin(n) \geq -1$，因此 $-1 <= sin(n), \forall n \geq 1$.
假设存在 u 满足 $u \leq x, \forall x \in \{\sin(n) : n \geq 1\}$ 且 $u > -1$
由 $\sin(\frac{3\pi}{2}) = -1 \in \{\sin(n) : n \geq 1\}$，得 $u \leq -1$，这与 $u > -1$ 矛盾，假设不成立.
所以 $\inf\{\sin(n) : n \geq 1\} = -1$.

Q2: Prove Cauchy-Schwartz inequality $\mathbf{x}^T\mathbf{y} \leq \|\mathbf{x}\|_2\|\mathbf{y}\|_2$ in two ways

方法1：
$\mathbf{x}^T\mathbf{y} = \sum_i x_iy_i$, $\|\mathbf{x}\|_2\|\mathbf{y}\|_2 = \sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}$
要证 $\mathbf{x}^T\mathbf{y} \leq \|\mathbf{x}\|_2\|\mathbf{y}\|_2$
即证 $\sum_i x_iy_i \leq \sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}$
即证 $(\sum_i x_iy_i)^2 \leq \sum_i x_i^2 \sum_i y_i^2$
即证 $\sum_{i=1}^n x_i^2y_i^2 + \sum_{i=1}^n \sum_{j=i+1}^n 2x_iy_ix_jy_j \leq \sum_{i=1}^n x_i^2y_i^2 + \sum_{i=1}^n \sum_{j=i+1}^n (x_i^2y_j^2 + x_j^2y_i^2)$
即证 $\sum_{i=1}^n \sum_{j=i+1}^n 2x_iy_ix_jy_j \leq \sum_{i=1}^n \sum_{j=i+1}^n (x_i^2y_j^2 + x_j^2y_i^2)$
由于 $(x_iy_j - x_jy_i)^2 = (x_i^2y_j^2 + x_j^2y_i^2) - 2x_iy_ix_jy_j \geq 0$，有 $2x_iy_ix_jy_j \leq (x_i^2y_j^2 + x_j^2y_i^2), \forall i, j \in [1, n]$
因此 $\sum_{i=1}^n \sum_{j=i+1}^n 2x_iy_ix_jy_j \leq \sum_{i=1}^n \sum_{j=i+1}^n (x_i^2y_j^2 + x_j^2y_i^2)$，原命题得证.

方法2：
若 $\mathbf{x} = 0$ 或 $\mathbf{y} = 0$，$\mathbf{x}^T\mathbf{y} = \|\mathbf{x}\|_2\|\mathbf{y}\|_2 = 0$，不等式成立.
否则 $\cos\langle\mathbf{x}, \mathbf{y}\rangle = \frac{\mathbf{x}^T \cdot \mathbf{y}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2} \leq 1$，$\|\mathbf{x}\|_2\|\mathbf{y}\|_2 > 0$，得 $\mathbf{x}^T\mathbf{y} \leq \|\mathbf{x}\|_2\|\mathbf{y}\|_2$，不等式成立.
综上，不等式成立.

Q3: Let $\mathbf{A}$ be a matrix of size $m \times n$. Denote the range space of $\mathbf{A}$ as $R(\mathbf{A})$ and the null space of $\mathbf{A}$ as $N(\mathbf{A})$, respectively. Prove $R(\mathbf{A}) = N^{\perp}(\mathbf{A}^T)$

$R(\mathbf{A}) = \{\mathbf{Ax}|\mathbf{x} \in \mathbb{R}^n\}$
$R^{\perp}(\mathbf{A}) = \{\mathbf{y}|\mathbf{x}^T\mathbf{y} = 0, \forall \mathbf{x} \in R(\mathbf{A})\}$
$N(\mathbf{A}^T) = \{\mathbf{x}|\mathbf{A}^T\mathbf{x} = 0\}$
$N^{\perp}(\mathbf{A}^T) = \{\mathbf{y}|\mathbf{x}^T\mathbf{y} = 0, \forall \mathbf{x} \in N(\mathbf{A}^T)\}$

先证 $\forall \mathbf{u} \in R(\mathbf{A}), \mathbf{u} \in N^{\perp}(\mathbf{A}^T)$:
假设 $\mathbf{u} \in R(\mathbf{A})$，则 $\exists \mathbf{v} \in \mathbb{R}^n, \mathbf{Av} = \mathbf{u}$.
任取 $\mathbf{x} \in N(\mathbf{A}^T)$，有 $\mathbf{A}^T\mathbf{x} = 0$，$(\mathbf{A}^T\mathbf{x})^T = \mathbf{x}^T\mathbf{A} = 0$，$\mathbf{x}^T\mathbf{Av} = \mathbf{x}^T\mathbf{u} = 0$.
所以 $\mathbf{u} \in N^{\perp}(\mathbf{A}^T)$，结论得证.

再证 $\forall \mathbf{u} \in N^{\perp}(\mathbf{A}^T), \mathbf{u} \in R(\mathbf{A})$:
即证 $\forall \mathbf{u} \in R^{\perp}(\mathbf{A}), \mathbf{u} \in N(\mathbf{A}^T)$.
假设 $\mathbf{u} \in R^{\perp}(\mathbf{A})$，则 $\forall \mathbf{x} \in \mathbb{R}^n, (\mathbf{A}\mathbf{x})^T\mathbf{u} = 0$，即 $\forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x}^T(\mathbf{A}^T\mathbf{u}) = 0$.
因此 $\mathbf{A}^T\mathbf{u} = 0$，即 $\mathbf{u} \in N(\mathbf{A}^T)$，结论得证.

综上，$R(\mathbf{A}) = N^{\perp}(\mathbf{A}^T)$ 得证.

Q4: For any two matrices, prove $trace(\mathbf{AB}) = trace(\mathbf{BA})$

假设 $\mathbf{A} \in \mathbb{R}^{n\times m}$，$\mathbf{B} \in \mathbb{R}^{m\times n}$:
$trace(\mathbf{AB}) = \sum_{i=1}^{n}(\mathbf{AB})_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{m}\mathbf{A}_{ij}\mathbf{B}_{ji} = \sum_{j=1}^{m}\sum_{i=1}^{n}\mathbf{B}_{ji}\mathbf{A}_{ij} = \sum_{j=1}^{m}(\mathbf{BA})_{jj} = trace(\mathbf{BA})$

Q5: Prove a useful inequality: $\mathbf{A} \succeq 0 \Leftrightarrow \langle \mathbf{A}, \mathbf{B} \rangle \geq 0$ for all $\mathbf{B} \succeq 0$

假设 $A, B \in \mathbb{R}^{n\times n}$:

先证充分性 $\mathbf{A} \succeq 0 \Rightarrow \langle \mathbf{A}, \mathbf{B} \rangle \geq 0, \forall \mathbf{B} \succeq 0$:
特征值分解 $\mathbf{A} = \mathbf{P}^T\mathbf{CP}, \mathbf{B}^T = \mathbf{Q}^T\mathbf{DQ}$，其中 $\mathbf{P}, \mathbf{Q}$ 为单位正交阵
$\langle \mathbf{A}, \mathbf{B} \rangle = trace(\mathbf{AB}^T) = trace(\mathbf{P^TCPQ^TD^TQ}) = trace(\mathbf{CDP}^T\mathbf{PQ}^T\mathbf{Q}) = trace(\mathbf{CD}) \geq 0$

再证必要性 $\langle \mathbf{A}, \mathbf{B} \rangle \geq 0, \forall \mathbf{B} \succeq 0 \Rightarrow \mathbf{A} \succeq 0$:
假设 $\mathbf{A} \not\succeq 0$，则 $\exists \mathbf{u} \neq 0, \mathbf{u}^T\mathbf{A}\mathbf{u} < 0$
取 $\mathbf{B} = \mathbf{u}\mathbf{u}^T$，则 $\langle \mathbf{A}, \mathbf{B} \rangle = trace(\mathbf{AB}^T) = trace(\mathbf{A}\mathbf{u}\mathbf{u}^T) = \langle \mathbf{A}\mathbf{u}, \mathbf{u} \rangle = \mathbf{u}^T\mathbf{A}\mathbf{u} < 0$
与条件矛盾，假设不成立.

原命题得证.

Q6: Define $f(\mathbf{x}) \triangleq \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$. Compute $\nabla f(\mathbf{x})$ and $\nabla^2 f(\mathbf{x})$

记 $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n))^T$, $\mathbf{a}_i \in \mathbb{R}^{1\times n}$

$(\nabla f(\mathbf{x}))_k = \frac{\partial}{\partial x_k}((\mathbf{A}\mathbf{x})_1 - b_1)^2 + \frac{\partial}{\partial x_k}((\mathbf{A}\mathbf{x})_2 - b_2)^2 + \cdots + \frac{\partial}{\partial x_k}((\mathbf{A}\mathbf{x})_n - b_n)^2 = \frac{\partial}{\partial x_k}(\mathbf{a}_1\mathbf{x} - b_1)^2 + \frac{\partial}{\partial x_k}(\mathbf{a}_2\mathbf{x} - b_2)^2 + \cdots + \frac{\partial}{\partial x_k}(\mathbf{a}_n\mathbf{x} - b_n)^2 = 2\mathbf{a}_{1k}(\mathbf{a}_1\mathbf{x} - b_1) + 2\mathbf{a}_{2k}(\mathbf{a}_2\mathbf{x} - b_2) + \cdots + 2\mathbf{a}_{nk}(\mathbf{a}_n\mathbf{x} - b_n) = 2\mathbf{a}_{1k}(\mathbf{A}\mathbf{x} - \mathbf{b})_1 + 2\mathbf{a}_{2k}(\mathbf{A}\mathbf{x} - \mathbf{b})_2 + \cdots + 2\mathbf{a}_{nk}(\mathbf{A}\mathbf{x} - \mathbf{b})_n = \sum_i 2\mathbf{a}_{ik}(\mathbf{A}\mathbf{x} - \mathbf{b})_i$

$\therefore \nabla f(\mathbf{x}) = 2\mathbf{A}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$

$(\nabla^2 f(\mathbf{x}))_{ij} = \frac{\partial}{\partial x_j}2\mathbf{a}_{1i}(\mathbf{a}_1\mathbf{x} - b_1) + \frac{\partial}{\partial x_j}2\mathbf{a}_{2i}(\mathbf{a}_2\mathbf{x} - b_2) + \cdots + \frac{\partial}{\partial x_j}2\mathbf{a}_{ni}(\mathbf{a}_n\mathbf{x} - b_n) = \sum_k 2\mathbf{a}_{ki}\mathbf{a}_{kj}$

$\therefore \nabla^2 f(\mathbf{x}) = 2\mathbf{A}^T\mathbf{A}$

Q7: Define $f(\mathbf{x}) \triangleq \|\mathbf{A} - \mathbf{x}\mathbf{x}^T\|_F^2$. Compute $\nabla f(\mathbf{x})$ and $\nabla^2 f(\mathbf{x})$

假设 $\mathbf{x} \in \mathbb{R}^n$:

$(\nabla f(\mathbf{x}))_k = \frac{\partial}{\partial xk}\sum_{i,j}(A_{ij} - x_i x_j)^2 = \sum_{i,j} -2(A_{ij} - x_i x_j)\frac{\partial}{\partial xk} x_i x_j = -2(A_{kk} - x_k^2)2x_k - 2\sum_{i\neq k}(A_{ik} + A_{ki} - 2x_i x_k)x_i = -2\sum_i (A_{ik} + A_{ki} - 2x_i x_k)x_i = 4x_k\sum_i x_i^2 - 2\sum_i x_i A_{ik} - 2\sum_i x_i A_{ki}$

$\therefore \nabla f(\mathbf{x}) = 4\mathbf{x}\mathbf{x}^T\mathbf{x} - 2\mathbf{A}\mathbf{x} - 2\mathbf{A}^T\mathbf{x}$

$(\nabla^2 f(\mathbf{x}))_{ij} = -\frac{\partial}{\partial x_j} 2\sum_k (A_{ki} + A_{ik} - 2x_k x_i)x_k = \begin{cases} \sum_k 4x_k^2 + 8x_i^2 - 4A_{ii}, i = j \\ 8x_i x_j - 2A_{ij} - 2A_{ji}, i \neq j \end{cases}$

$\therefore \nabla^2 f(\mathbf{x}) = 8\mathbf{x}\mathbf{x}^T + 4\mathbf{x}^T\mathbf{x}\mathbf{I} - 2\mathbf{A} - 2\mathbf{A}^T$

Q8: For the logistic regression example in lecture notes, compute $\nabla^2 E(\mathbf{w})$

Q9: Define $f(\mathbf{x}) \triangleq \log\sum_{k=1}^n(\exp(x_k))$. Prove $\nabla^2 f(\mathbf{x}) \succeq 0$

Q10: Find at least one example in either of the following two fields that can be formulated as an optimization problem and show how to formulate it: 1.EDA software 2.cluster scheduling for data centers