24 Fall

Optimization Methods for Analytics

"cheat sheet"

Linear Algebra

* Notation

- matrix 矩陣 : $A \in \mathbb{R}^{m \times n}$ (rows × columns)
- vector 向量 : $b \in \mathbb{R}^n$ → ① points ② directions
- element 元素 : x_i
- $\mathbb{R}$ 實數, $\mathbb{Z}$ 整數

* Matrix operations

1. Sum $C_{ij} = A_{ij} + B_{ij}$ (dimensions need to be the same)

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} + \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11}+B_{11} & A_{12}+B_{12} \\ A_{21}+B_{21} & A_{22}+B_{22} \end{pmatrix}$$

2. transpose $A_{ij}{}^T = A_{ji}$

$$\underbrace{A = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix} \rightarrow A^T = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}}_{\text{symmetric matrix}}, \quad B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \rightarrow B^T = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \rightarrow b^T = (1 \ 2 \ 3)$$

3. product $A^{m \times n} \cdot B^{n \times p} = C^{m \times p}$ $\quad C = A \cdot B \neq BA$

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad B = \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix} \rightarrow C = AB = \begin{pmatrix} 1\times0+2\times2 & 1\times1+2\times3 \\ 3\times0+4\times2 & 4\times1+4\times3 \end{pmatrix} = \begin{pmatrix} 4 & 7 \\ 8 & 15 \end{pmatrix}$$

I : identity → $A = AI = IA$ square matrix

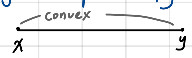
4. other

- $A(BC) = (AB)C$
- $A(B+C) = AB + AC$

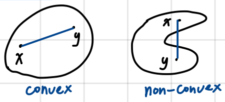
Convexity

* What is convex 凸函數

- **Convex combination** 凸(線性)組合 : give two points $x, y \in \mathbb{R}^n$, the point z is a convex combination of x and y, if $z = \lambda x + (1-\lambda) y$, $0 \le \lambda \le 1$



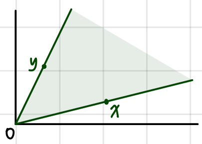
- **Convex set** 凸集合 : a set $S \subset \mathbb{R}^n$ is convex, if $x, y \in S \rightarrow \lambda \cdot x + (1-\lambda) \cdot y \in S$ $\quad \forall \lambda \in [0,1]$



- cone : 在集合S裡任一點x放大λ倍後仍在集合裡 → S為 convex cone

$x \in S \rightarrow \lambda x \in S \quad \forall \lambda \ge 0$

- **conic combination** : give two points $x, y \in \mathbb{R}^n$, the point z is a conic combination of x and y, if $z = \lambda_1 \cdot x + \lambda_2 \cdot y$, $\lambda_1 . \lambda_2 \ge 0$

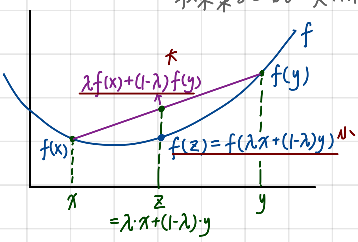


[convex : 有限制係數總合為1
[conic : 各自放大無限制

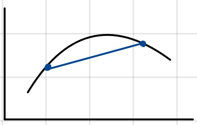
- **convex function** 凸函數 : give a convex set $S \subseteq \mathbb{R}^n$, a function $f : S \rightarrow \mathbb{R}$ is a convex function

if $f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y)$ $\quad \forall x . y \in S$, $\forall \lambda \in [0,1]$ 包括

取某集合S的2點x和y, 對於所有x.y連線上的點都不小於函數本身, 此函數 is convex function



- strictly convex function : $f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y)$ $\quad \forall x . y \in S$, $\forall \lambda \in (0,1)$ 不包括
- Concave function : $f(\lambda x + (1-\lambda) y) \ge \lambda f(x) + (1-\lambda) f(y)$ $\quad \forall x, y \in S$, $\forall \lambda \in [0,1]$
  - △ function f is concave if and only $-f$ is convex

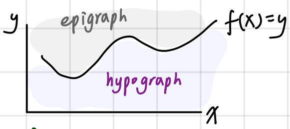


* convex 的性質

- **epigraph** : 函數之上的元素集合

$epi(f) = \{ (x,y) \in \mathbb{R}^{n+1} : f(x) \le y \}$

△ function f is convex if and only $epi(f)$ is convex



- **hypograph** : 函數之下的元素集合

$hyp(f) = \{ (x,y) \in \mathbb{R}^{n+1} : f(x) \ge y \}$

△ function f is concave if and only $hyp(f)$ is convex

* How to recognize convexity

- let $f : \mathbb{R} \rightarrow \mathbb{R}$ be convex

then $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ define as : $g(x) = f(a^T x)$ is convex

$g\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = e^{-x_1 + 5x_2}$, $\underbrace{f(x) = e^x}_{\text{convex}}$, $f'(x) = e^x$, $f''(x) = e^x \ge 0$

$\rightarrow g = f(-x_1 + 5x_2)$

$f(x)$ is convex, $g(x) = f(a^T \cdot x)$ is also convex

x次方的微分
- $f(x) = 2^x$
- $\rightarrow f'(x) = 2^x \cdot \ln 2 = 2^x$
- , 同理 $g(x) = e^x$
- $\rightarrow g'(x) = e^x \cdot \ln e = e^x$

- if $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}$ are convex,

then $h(x) = \lambda_1 \cdot f(x) + \lambda_2 \cdot g(x)$, with $\lambda_1, \lambda_2 \ge 0$

$h(x)$ is convex

$f(x)$ and $g(x)$ is convex → $f(x) + g(x)$ is also convex

e.g.

$f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = e^{-x_1 + 5x_2} + x_1^2 + x_2^4$

$f : \mathbb{R} \rightarrow \mathbb{R}$ → if $f''\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \ge 0$ → is convex

$f(x_1) = x_1^2$, $f'(x_1) = 2x_1$, $f''(x_1) = 2 \ge 0$

$f(x_2) = x_2^4$, $f'(x_2) = 4x_2^3$, $f''(x_2) = 12x_2^2 \ge 0$

* convexity and optimization

- global minimum

S is a non-empty set, $x \in S$ is a global minimum of $\min_{x \in S} f(x)$
if $f(x) \leq f(\bar{y})$ for all $\bar{y} \in S$

- local minimum

S is a non-empty convex set, $x \in S$ is a local minimum of $\min_{x \in S} f(x)$
if $f(x) \leq f(\bar{y})$ for all $\bar{y} \in S \cap \{y \in \mathbb{R}^n : \|y - x\|_2 \leq \epsilon\}$
when $\epsilon > 0$ exists

- convex optimization

① if S is convex set, f is convex function $\rightarrow$ local minimum of P are also global minimum
$\Rightarrow P$ is convex optimization problem

② if S is convex set, f is concave function $\rightarrow$ P has optimal solution = extreme point

* Proof of convexity

① $S_1 = \{x \in \mathbb{R}^n : \max\{x_1, x_2, \cdots, x_n\} \leq 1\}$ is convex

$\rightarrow S_1 = \underbrace{\{x \in \mathbb{R}^n : x_1 \leq 1\}}_{\text{convex}} \cap \underbrace{\{x \in \mathbb{R}^n : x_2 \leq 1\}}_{\text{convex}} \cap \cdots \underbrace{\{x \in \mathbb{R}^n : x_n \leq 1\}}_{\text{convex}}$

$\rightarrow S_1$ is convex $\because$ it is the iteration of convex set

② $S_1 = \{x \in \mathbb{R}^n : \sum_{i=1}^n |x_i| \leq 1\}$ is convex

$\rightarrow S_1 = \{x \in \mathbb{R}^n : g(x) \leq 1\}$ where $g(x) = \underbrace{\sum_{i=1}^n \underbrace{|x_i|}_{\text{convex}}}_{\text{convex}}$

let $x, y \in S$, $0 \leq \lambda \leq 1$ (we need to show that $z = \lambda x + (1-\lambda) y \in S$)

$\because x, y \in S \rightarrow g(x) \leq 1$, $g(y) \leq 1$

$g(\lambda x + (1-\lambda) y) \underset{\text{convex}}{\leq} \lambda g(x) + (1-\lambda) g(y) \leq \max\{g(x), g(y)\}$

a) Given two convex functions $f, g : \mathbb{R}^n \to \mathbb{R}$, and scalars $\lambda_1, \lambda_2 \geq 0$, prove that the weighted sum $h : \mathbb{R}^n \to \mathbb{R}$ defined as $h(\boldsymbol{x}) = \lambda_1 f(\boldsymbol{x}) + \lambda_2 g(\boldsymbol{x})$ is convex.

Since $f, g : \mathbb{R}^n \to \mathbb{R}$ are all convex functions, by definition, we have that

$$\forall x_1, x_2 \in \mathbb{R}^n \text{ and } \forall \mu_1 \in [0,1], f[\mu_1 x_1 + (1-\mu_1) x_2] \leq \mu_1 f(x_1) + (1-\mu_1) f(x_2)$$
$$\forall y_1, y_2 \in \mathbb{R}^n \text{ and } \forall \mu_2 \in [0,1], g[\mu_2 y_1 + (1-\mu_2) y_2] \leq \mu_2 g(y_1) + (1-\mu_2) g(y_2)$$

Given $\lambda_1, \lambda_2 \geq 0$, we have $\forall z_1, z_2 \in \mathbb{R}^n$ and $\forall \mu \in [0,1]$,

$$\begin{aligned} h[\mu z_1 + (1-\mu) z_2] &= \lambda_1 f[\mu z_1 + (1-\mu) z_2] + \lambda_2 g[\mu z_1 + (1-\mu) z_2] \\ (f, g \text{ convex}; \lambda_1, \lambda_2 \geq 0) &\leq \lambda_1 [\mu f(z_1) + (1-\mu) f(z_2)] + \lambda_2 [\mu g(z_1) + (1-\mu) g(z_2)] \\ &= \mu [\lambda_1 f(z_1) + \lambda_2 g(z_1)] + (1-\mu) [\lambda_1 f(z_2) + \lambda_2 g(z_2)] \\ &= \mu h(z_1) + (1-\mu) h(z_2) \end{aligned}$$

We can see that h satisfies the definition of a convex function, and is thus convex.

b) Given two convex sets $A, B \subseteq \mathbb{R}^n$, prove that the intersection $A \cap B = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} \in A \text{ and } \boldsymbol{x} \in B\}$ is convex.

Suppose $x, y \in A \cap B$, then we have $x, y \in A$ and $x, y \in B$. Since the sets A and B are convex, by definition, we have $\forall \lambda \in [0,1]$, $\lambda x + (1-\lambda) y \in A$ and $\lambda x + (1-\lambda) y \in B$, which means that $\lambda x + (1-\lambda) y \in A \cap B$. Thus, we can conclude that the set $A \cap B$ is also convex.

c) Given a convex set $A \subseteq \mathbb{R}^n \times \mathbb{R}^m$, prove that the projection $\text{proj}(A) = \{\boldsymbol{x} \in \mathbb{R}^n : \exists \boldsymbol{y} \in \mathbb{R}^m \text{ such that } (\boldsymbol{x}, \boldsymbol{y}) \in A\}$ is convex.

Suppose $x_1 \in \text{proj}(A) = \{x_1 \in \mathbb{R}^n : \exists y_1 \in \mathbb{R}^m \text{ such that } (x_1, y_1) \in A\}$ and $x_2 \in \text{proj}(A) = \{x_2 \in \mathbb{R}^n : \exists y_2 \in \mathbb{R}^m \text{ such that } (x_2, y_2) \in A\}$. To show $\text{proj}(A)$ is a convex set, we need to show $\lambda x_1 + (1-\lambda) x_2 \in \text{proj}(A)$, namely

$$\exists y \in \mathbb{R}^m \text{ such that } [\lambda x_1 + (1-\lambda) x_2, y] \in A$$

To find such y, we set $y = \lambda y_1 + (1-\lambda) y_2$, then by convexity of the set A we have

$$[\lambda x_1 + (1-\lambda) x_2, \lambda y_1 + (1-\lambda) y_2] \in A$$

The proof is complete.

Formulation

* Optimization formulation

$$\underset{x \in \mathbb{R}^n}{\overset{(\max)}{\min}} f(x) \leftarrow \text{objective function}$$

decision variable $\longrightarrow x \in \mathbb{R}^n$

subject to $\rightarrow$ s.t. $x \in S \leftarrow$ Feasible region (constraints)

* Integer (binary) Optimization

$$\underset{\min}{\max} \sum_{i \in A} b_i x_i$$
$$\text{s.t. } \sum_{i \in A} w_i x_i \leq w_{max, min}$$
$$x_i \in \{0, 1\} \quad \forall i \in A$$

* Linear Optimization

$$\underset{\min}{\max} \sum_{i \in A} b_i x_i$$
$$\text{s.t. } \sum_{i \in A} w_i x_i \leq w_{max, min}$$
$$0 \leq x_i \leq 1 \quad \forall i \in A$$

Linear Regression

LP	$\min\ c^T x$ s.t. $Ax \leq b$	NLP	$\min\ f(x)$ s.t. $x \in S$
MIP	$\min\ c^T x$ s.t. $Ax \leq b$ $x \in \mathbb{Z}$	convex	$\min\ f(x)$ s.t. $x \in S$

* Linear Regression

- Predictors : $a_i = (a_{i1}, a_{i2}, \cdots, a_{in})_{i=1}^m$ n-dimensional vectors
- Response variable : $(y_i)_{i=1}^m$

- assumptions : $y_i = \underbrace{x_0}_{\text{① Intercept}} + \sum_{j=1}^n a_{ij} \underbrace{x_j}_{\text{② coefficient}} + \underbrace{\varepsilon_i}_{\text{③ remainder}}$

* Ordinary Least Squares (OLS)

solution :
$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^n \left(y_i - x_0 - \sum_{j=1}^n a_{ij} x_j \right)^2$$

- shortcomings of OLS $\rightarrow$ when num of predictors is large.
  - $\rightarrow$ sensitive to outliers, overfitting, hard to interpret
  - $\rightarrow$ poor generalizability : $R^2 \uparrow$, train data perfect but can't predict
  - $\rightarrow$ does not identify critical predictors : always use all predictor variables
  - $\rightarrow$ unless prediction > observation

* Parsimony (simplest are often right)

- Best subset selection $\rightarrow$ expensive computation

Nonlinear Optimization

*Gradient 梯度,斜率

- let $f: \mathbb{R}^n \to \mathbb{R}$, gradient of f at x = $f'(x)$ or $\nabla f(x)$

$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}$

一維度:純量x的梯度 $f'(x) \to f(x)$對x的微分
多維度:向量x的梯度 $\nabla f(X) \to f(x)$對x所有元素的微分

$f(x) = x_1^2 - 2x_1x_2 + 2x_2^2$ find gradient of $f(x)$

$f'(x_1) = 2x_1 - 2x_2 \quad f'(x_2) = 4x_2 - 2x_1$

$\nabla f(x) = \begin{pmatrix} 2x_1 - 2x_2 \\ 4x_2 - 2x_1 \end{pmatrix}$

$\nabla f\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 4 \end{pmatrix}, \nabla f\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \nabla f\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

- let $f: \mathbb{R}^n \to \mathbb{R}$, directional derivative of f at x, in direction d $\to$ $f'(x;d) = \lim_{\lambda \to 0} \frac{f(x+\lambda d) - f(x)}{\lambda}$

$f'(x;d) = \nabla f(x)^T d$

*Necessary and sufficient condition

- stationary point: points that satisfy the property of gradient
  $f: \mathbb{R} \to \mathbb{R}$ 情況下的臨界点, when derivate $=0$, 從又小的x不会 inf y
- Necessary condition: $\bar{x}$ is a local minimum, $\nabla f(x) = 0$
- sufficient condition: if $f(x)$ is convex, $\nabla f(x) = 0 \to \bar{x}$ is a global minimum

*Find stationary points

- Unconstrained quadratic optimization
  $\to$ only used in quadratic optimization problem (二次方程式)

$f(x) = (2 - x_1 - x_2)^2 + (3 - 2x_1)^2 + (2 - x_1 - 2x_2)^2$

$\frac{\partial f}{\partial x_1}(x) = -2(2 - x_1 - x_2) - 4(3 - 2x_1) + 2(2 - x_1 - 2x_2) = 12x_1 + 6x_2 - 20$

$\frac{\partial f}{\partial x_2}(x) = -2(2 - x_1 - x_2) - 4(2 - x_1 - 2x_2) = 6x_1 + 10x_2 - 12$

$\begin{cases} 12x_1 + 6x_2 - 20 = 0 \\ 6x_1 + 10x_2 - 12 = 0 \end{cases} \to \begin{cases} x_1 = \frac{2}{9} \\ x_2 = \frac{32}{21} \end{cases}$

- Descent method
  1. iteration $k=0$, solution x^0
  2. If $\nabla f(x^k) \approx 0$, stop and return x^k
  3. find direction $d^k \in \mathbb{R}^n$
  4. find length (step size) $\lambda^k > 0$, $f(x^k + \lambda^k d^k) < f(x^k)$
  5. set $x^{k+1} \leftarrow x^k + \lambda^k d^k$ and $k \leftarrow k+1$, go to 2

- Line Search
  - Binary Search
    1. $a < b$ and $f(a)f(b) < 0$, set an interval $\in$ $\to$ without loss of generality
    2. $c = (a+b)/2$
    3. if $f(c) \le 0$, let $a \leftarrow c$; else if $f(c) \ge 0$ let $b \leftarrow c$
    4. if $b - a < \epsilon$, stop, else go to 2 and repeat
       $\Rightarrow$ exist $x \in [a,b]$
  - Newton method
    ① Find equation $g(x) = 0$

    $$x^{k+1} = \boxed{x^k - \frac{g(x^k)}{g'(x^k)}}$$

    ② Find minimum $g(x)$

    $$x^{k+1} = \boxed{x^k - \frac{g'(x^k)}{g''(x^k)}}$$

[Equation]

Find a solution of $x^2 - 313 = 0$

① $f(x) = x^2 - 313$

$a=0, b=20, c=10, f(c) = 100 - 313 < 0$

$a=10, b=20, c=15, f(c) = 225 - 313 < 0$

$a=15, b=20, c=17.5 \ \cdots$

② $f(x) = x^2 - 313 = 0$

$f'(x) = 2x$

set initial $x^0 = 20$

$x^{k+1} = 20 - \frac{400 - 313}{20 \times 2} = 17.\sim$

$\vdots$

[minimum]

Find minimum of $f(x) = 2x^3 - e^x$

$=$ Find solution of $g(x) = 6x^2 - e^x = 0$

$f(x) = \frac{2}{3}|x|^{\frac{3}{2}}$

$\begin{cases} x > 0 \to f(x) = \frac{2}{3}x^{\frac{3}{2}} \\ \quad f'(x) = x^{\frac{1}{2}} \\ \quad f''(x) = \frac{1}{2}x^{-\frac{1}{2}} \\ x < 0 \to f(x) = \frac{2}{3}(-x)^{\frac{3}{2}} \\ \quad f'(x) = -(-x)^{\frac{1}{2}} \\ \quad f''(x) = \frac{1}{2}(-x)^{-\frac{1}{2}} \end{cases}$

Question 1 (LP)

A farmer has 100 acres of land on which she plans to grow wheat and corn. Each acre of wheat requires 4 hours of labor and $20 of capital, and each acre of corn requires 16 hours of labor and $40 of capital. The farmer has at most 800 hours of labor and $2400 of capital available. If the profit from an acre of wheat is 80, and from an acre of corn is 100, how many acres of each crop should she plant to maximize her profit? Formulate a *linear program* that solves this problem (no need to actually solve it)

decision variable:

X_w, X_c the number of arce planting wheat and corn

$\max\ 80X_w + 100X_c$

s.t. $X_w + X_c \le 100$ (land)

$4X_w + 16X_c \le 800$ (labor)

$20X_w + 40X_c \le 2400$ (capital)

Question 2 (LP)

During each 6-hour period of the day, the Los Angeles Police Department needs at least the number of policeman shown below. Policeman can be hired to work either 12 consecutive hours or 18 consecutive hours. Policeman are paid $10 per hour for each of the first 12 hours a day they work, and are paid $15 per hour for the next 6 hours they work in a day. Formulate a *linear program* that can be used to minimize the cost of meeting LA's daily police requirements.

Time period	Number of policemen required
12AM-6AM	12
6AM-12PM	8
12PM-6PM	6
6PM-12AM	15

decision variable: $X_{ab}, X_{bc}, X_{cd}, X_{da}$ the number of police in each period

$X_{abc}, X_{bcd}, X_{cda}, X_{dab}$

$\min\ 10\times12\times(X_{ab}+X_{bc}+X_{cd}+X_{da}) + (10\times12+15\times6)(X_{abc}+X_{bcd}+X_{cda}+X_{dab})$

s.t. $X_{ab}+X_{da}+X_{abc}+X_{cda}+X_{dab} \ge 12$ (1st period)

$X_{ab}+X_{bc}+X_{abc}+X_{bcd}+X_{dab} \ge 8$ (2nd period)

$X_{bc}+X_{cd}+X_{abc}+X_{bcd}+X_{cda} \ge 6$ (3rd period)

$X_{cd}+X_{da}+X_{bcd}+X_{cda}+X_{dab} \ge 15$ (4th period)

Question 3 (LP)

A firm offering cloud computing services needs to decide how to allocate jobs to available servers optimally. In this particular example there are four jobs to process, and the time required to complete each job is given in Table 1. Additionally, the firm has three servers available . All servers have different characteristics, and therefore not all servers are suitable to process each job; specifically, Table 2 shows which servers are compatible with each job. A job can be split among two different compatible servers and processed independently. The objective is to minimize the time required to complete all jobs: since all jobs are processed in parallel, this is time used by the server that finishes processing jobs last.

Table 1: Time required to complete each job (in hours)

Job	1	2	3	4
Hours	100	150	400	200

Table 2: Compatibility of jobs and servers

Job	1	2	3	4
Server 1	x		x	
Server 2		x	x	x
Server 3	x	x		x

As an example, if Job 1 is processed completely in server 1, Job 2 is processed completely in Server 2, Job 3 is split evenly between Server 1 and Server 2 and Job 4 is processed completely in Server 3, then:

- Server 1 works for a total of $100 + \frac{1}{2}400 = 300$ hours.
- Server 2 works for a total of $150 + \frac{1}{2}400 = 350$ hours.
- Server 3 works for a total of 200 hours.
- The time required to complete all jobs (in parallel) is 350 hours.

Formulate the problem above as an *linear program.* Clearly define your notation, variables and include a short description of each constraint.

decision variable: X_{ij} the propotion of Job I assign to Server J

$I = \{1,2,3,4\}$, $\forall i \in I$

$J = \{1,2,3\}$, $\forall j \in J$

T is the maximum working hour

$\min\ T$

s.t. $X_{11} + X_{13} = 1$ (Job 1)

$X_{22} + X_{23} = 1$ (Job 2)

$X_{31} + X_{32} = 1$ (Job 3)

$X_{42} + X_{43} = 1$ (Job 4)

$100X_{11} + 400X_{31} \le T$ (server 1)

$150X_{22} + 400X_{32} + 200X_{42} \le T$ (server 2)

$100X_{13} + 150X_{23} + 200X_{43} \le T$ (server 3)

$X_{ij} \ge 0$

Question 4 (MIP)

Coach Billy Beane and his assistant Peter Brand have been hired by the Lakers. He is trying to choose the starting lineup for the basketball team. The team consists of seven players who have been rated (on a scale of 1= poor to 3 = excellent) according to their ball-handling, shooting, rebounding, and defensive abilities. The positions that each player is allowed to play and the player's abilities are listed below.

Player	Position	Ball-handling	Shooting	Rebounding	Defense
1	G	3	3	1	3
2	C	2	1	3	2
3	G-F	2	3	2	2
4	F-C	1	3	3	1
5	G-F	3	3	3	3
6	F-C	3	1	2	3
7	G-F	3	2	2	1

The five-player starting lineup must satisfy the following restrictions:

1. At least 4 members must be able to play guard, at least two members must be able to play forward, and at least 1 member must be able to play center.
2. The average ball-handling, shooting, and rebounding level of the starting lineup must be at least 2.
3. If player 3 starts, then player 6 cannot start.
4. If player 1 starts, then players 4 and 5 must both start.
5. Either player 2 or player 3 must start.

Given those constraints, Coach Beane wants to maximize the total defensive ability of the starting team. Formulate a mixed-integer program that will help him choose his starting team.

decision variable: $X_i \begin{cases} 1 & \text{chosen} \\ 0 & \text{not chosen} \end{cases}$ of player $i \in (1,2,3,4,5,6,7)$

max $3X_1 + 2X_2 + 2X_3 + X_4 + 3X_5 + 3X_6 + X_7$

s.t. $\sum_{i=1}^{7} X_i = 5$ (players)

$X_1 + X_3 + X_5 + X_7 \geq 4$ (guard)

$X_3 + X_5 + X_7 \geq 2$ (front)

$X_2 + X_4 + X_6 \geq 1$ (center)

$\frac{1}{7}(3X_1 + 2X_2 + 2X_3 \cdots) \geq 2$ (ball-handling)

$\frac{1}{7}(3X_1 + X_2 + 3X_3 \cdots) \geq 2$ (shooting)

$\frac{1}{7}(X_1 + 3X_2 + 2X_3 \cdots) \geq 2$ (rebounding)

Question 5 (MIP)

You are tasked with planning which movies to reference in the remainder of this course. You have a list of n movies, and for each movie you have the basic information: budget, box office gross earnings, release dates, gender of the main protagonist, film rating, Rotten Tomatoes score, and number of academy awards won. Finally, making a reference to a movie requires a time investment (for example, seeing the movie). In other words, you have access to the following table.

#	Name	Budget	Earnings	Date	Gender	Rating	Score	Awards	Time
1	Terminator 2: Judgment Day	\$102M	\$520M	1991	M	R	93%	4	137min
2	Star Wars: The Force Awakens	\$306M	\$2068M	2015	F	PG-13	93%	0	138min
3	LOTR: Return of the King	\$94M	\$1146M	2003	M	PG-13	93%	11	∞
$\vdots$	$\vdots$	$\ddots$							
n	$\dots$	$\dots$							

So far, in the course, movies have been selected in order to maximize the earnings (adjusted by inflation) of the movies referenced, while ensuring a reasonable time commitment. In other words, letting r_i be the earnings of movie i (adjusted by inflation), letting t_i be the time investment required to reference movie i, and letting T be the total time available to watch movies, the movies selected correspond to the optimal solutions of the optimization problem

$$\max_x \sum_{i=1}^{n} r_i x_i \quad \text{(earnings)}$$

$$\text{s.t} \sum_{i=1}^{n} t_i x_i \leq T \quad \text{(time constraint)}$$

$$x \in \{0,1\}^n.$$

In the formulation above, x_i is a decision variable such that $x_i = 1$ if movie i is selected to reference in class, and $x_i = 0$ otherwise.

Do the following modifications to the above formulation to ensure better movies are chosen. Clearly define the notation you introduce, as well as additional decision variables (if needed).

1. You have noticed that, so far, all references used involve white males. Modify the above formulation so that at least 30% of the movies used have female leads.
2. The movies should have some intellectual merit as well. Modify the formulation so that the average number of awards per movie is at least 2.
3. It may be OK to have some "mature" references, but the class should not become too dark. Modify the formulation so that at most one R rated movie is used.
4. Since high-earning movies are preferred, there is also an implicit bias towards high-budget films. Modify the formulation so that for each movie referenced with a budget over \$100M, there should also be a movie with a budget under \$10M.

Question 1

The university of Southern California offers the following seven courses in the PhD program in a given year. The name of the courses, credits required and utility (if taken) is described in the following table.

#	Name	Credits	Utility
1	Linear optimization	4	5
2	Stochastic processes	4	3
3	Large-scale optimization	4	4
4	Quantum computing	4	7
5	Decision analysis	2	2
6	Department seminar	1	2
7	Directed research	1 (can be taken multiple times)	1

Suppose you can take up to 14 units among the offered courses. The problem of choosing the best selection of courses to maximize the utility is given by

$$\max_x 5x_1 + 3x_2 + 4x_3 + 7x_4 + 2x_5 + 2x_6 + x_7 \quad \text{(Utility)}$$

$$\text{s.t. } 4(x_1 + x_2 + x_3 + x_4) + 2x_5 + x_6 + x_7 \leq 14 \quad \text{(Credit hours)}$$

$$x_i \in \{0,1\} \quad \forall i = 1,\dots,6 \quad \text{(Binary variables)}$$

$$x_7 \in \mathbb{Z}, \quad \text{(Directed research can be taken multiple times)}$$

where x_i indicates whether a course if taken or not (if $1 \leq i \leq 6$) or the number of credits taken (if $i = 7$).

Modify the formulation to account for the following considerations.

1. Quantum computing and decision analysis are scheduled at the same time, so it is not possible to take both courses.
2. Linear optimization is a prerequisite of large-scale optimization. Thus, to take the later course, linear optimization must also be taken.
3. To ensure breadth of the program, it is not possible to take all courses with "optimization" or "computing" in the name.
4. Students who do not undertake any units of directed research must be enrolled in the department seminar.
5. Students are required to take at least two of the following three fundamental courses: linear optimization, stochastic processes and decision analysis.

*Final
- steepest descent + Newton method
- compute dual
- Given primal and dual, and given optimal x^*/y^*, find the other using CS
- modeling 1
- modeling 2

$\min c^\top x$	$\max y^\top b$
$a_i^\top x \geq b_i$	$y_i \geq 0$
$a_i^\top x \leq b_i$	$y_i \leq 0$
$a_i^\top x = b_i$	y_i free
$x_j \geq 0$	$y^\top A_j \leq c_j$
$x_j \leq 0$	$y^\top A_j \geq c_j$
x_j free	$y^\top A_j = c_j$