1. From the hint that y_j has a Poisson distribution with mean $\sum_{i=1}^{n} p_{ji}\mu_i = p_j^T\mu$

Thus, $\log(\Pr(y_j=k)) = -p_j^T\mu + k\log(p_j^T\mu) - \log k!$

Suppose $k_1,\dots,k_n$ are observed value of y_j. Then the estimation problem is

$$\begin{aligned} \max \quad & -\sum_{j=1}^{m} p_j^T\mu + \sum_{j=1}^{m} k_j \log(p_j^T\mu) \\ \text{s.t.} \quad & \mu \succeq 0 \end{aligned}$$

which is a convex optimization problem in μ.

2. The weight error margin can be expressed as the maximal ε such that

$$(a+\eta)^T x_i \geq b, \quad i=1,\dots,N$$
$$(a+\eta)^T y_i \leq b, \quad i=1,\dots,M$$

for all $\|\eta\|_2 \leq \varepsilon$.

This shows that the weight error margin is given by

$$\min_{\substack{i=1,\dots,N \\ j=1,\dots,M}} \left(\frac{a^T x_i - b}{\|x_i\|_2}, \frac{b-a^T y_j}{\|y_j\|_2} \right)$$

which can be reformulated as

$$\begin{aligned} \max \quad & t \\ \text{s.t.} \quad & a^T x_i - b \geq t\|x_i\|_2, \quad i=1,\dots,N \\ & b - a^T y_j \geq t\|y_j\|_2, \quad j=1,\dots,M \\ & \|a\|_2 \leq 1 \end{aligned}$$

3. $\min \quad t$

$$\text{s.t.} \quad x_i^T P x_i + q^T x_i + r \geq 0, \quad i = 1, \ldots, N$$
$$y_j^T P y_j + q^T y_j + r \leq 0, \quad j = 1, \ldots, M$$
$$I \preceq P \preceq tI$$

which is an SPP with optimization variable $P \in S^n$, $q \in R^n$, $r \in R$, $t \in R$

4. From the definition of m-strongly convex, we have

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{m}{2} \|y-x\|^2$$

Similarly, we can have

$$f(x) \geq f(y) + \nabla f(y)^T (x-y) + \frac{m}{2} \|x-y\|^2$$

Adding both inequalities gives

$$f(y) + f(x) \geq f(x) + f(y) + (x-y)^T (\nabla f(y) - \nabla f(x)) + m\|x-y\|^2$$

which is simplified to

$$(\nabla f(x) - \nabla f(y))^T (x-y) \geq m\|x-y\|^2$$

5. (a) Let $\bar{x} = \alpha x_1 + (1-\alpha) x_2$, $\alpha \in [0,1]$. From the definition of M-smooth function, we have

$$f(x_1) \leq f(\bar{x}) + \nabla f(\bar{x})^T (x_1 - \bar{x}) + \frac{M}{2} \|x_1 - \bar{x}\|^2$$
$$f(x_2) \leq f(\bar{x}) + \nabla f(\bar{x})^T (x_2 - \bar{x}) + \frac{M}{2} \|x_2 - \bar{x}\|^2$$

$$\Rightarrow \begin{array}{l} \alpha f(x_1) \leq \alpha f(\bar{x}) + \alpha \nabla f(\bar{x})^T (x_1 - \bar{x}) + \alpha \frac{M}{2} \|x_1 - \bar{x}\|^2 \\ (1-\alpha) f(x_2) \leq (1-\alpha) f(\bar{x}) + (1-\alpha) \nabla f(\bar{x})^T (x_2 - \bar{x}) + (1-\alpha) \frac{M}{2} \|x_2 - \bar{x}\|^2 \end{array}$$

$$\Rightarrow \alpha f(x_1) + (1-\alpha) f(x_2) \leq f(\bar{x}) + \nabla f(\bar{x})^T (\alpha x_1 + (1-\alpha) x_2 - \bar{x}) + \frac{M}{2} (\alpha(1-\alpha) \|x_1 - x_2\|^2)$$

$$\Rightarrow f(\bar{x}) = f(\alpha x_1 + (1-\alpha) x_2) \geq \alpha f(x_1) + (1-\alpha) f(x_2) - \frac{\alpha(1-\alpha) M}{2} \|x_1 - x_2\|^2$$

(b). Similarly to problem 4, we can have

$$\begin{array}{l} f(x) \leq f(y) + \nabla f(y)^T (x-y) + \frac{M}{2} \|x-y\|^2 \\ f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{M}{2} \|x-y\|^2 \end{array} \Rightarrow (\nabla f(x) - \nabla f(y))^T (x-y) \leq M\|x-y\|^2$$