2a) Learn an unknown positive semidefinite matrix

$$X^* = U^* U^{*T}$$

where $U^* \in \mathbb{R}^{d \times r'}$ $\Rightarrow$ rank - r matrix.

Given:

- m input samples,
  - ↳ linear measurements of X^*

For each sample $i = 1, 2, \ldots, m$,

let $A_i \in \mathbb{R}^{d \times d}$ be a random matrix

where every A_i is drawn independently from Guassian distribution, mean = 0, variance = 1.

$$y_i = \langle A_i, X^* \rangle$$

Solve (or) recover X^*.

Let $U \in \mathbb{R}^{d \times R}$ be the variable matrix.

Minimizing following mean squared loss;

$$f(U) = \frac{1}{2m} \sum_{i=1}^{m} (\langle A_i, UU^T \rangle - y_i)^2,$$

compute gradient of $f(U)$ over U.

Solution:

Minimizing $f(U)$ is nothing but optimization by altering U.

Taking derivatives,

$$f(U) = (\langle A_i, UU^T \rangle - y_i)^2 \quad \text{[removing constants]} \qquad (1)$$

Adding $\underset{\text{smaller term.}}{V}$ on both sides, to bring taylor expansion term,

$$f(U+V) = (\langle A_i, (U+V)(U+V)^T \rangle - y_i)^2 \qquad (2)$$

Subtracting (1) from (2).

$$f(U+V) - f(U) = (\langle A_i, (U+V)(U+V)^T \rangle - y_i)^2 - (\langle A_i, UU^T \rangle - y_i)^2$$

$$\Rightarrow (\langle A_i, (U+V)(U+V)^T \rangle)^2 - 2(\langle A_i, (U+V)(U+V)^T \rangle) y_i + y_i^2$$
$$- [(\langle A_i, UU^T \rangle)^2 - 2 \langle A_i, UU^T \rangle y_i + y_i^2]$$

$$\Rightarrow (\langle A_i, (U+V)(U+V)^T \rangle)^2 - 2 y_i \langle A_i, (U+V)(U+V)^T \rangle + \cancel{y_i^2}$$
$$- \langle A_i, UU^T \rangle^2 + 2 \langle A_i, UU^T \rangle y_i - \cancel{y_i^2}$$

Scanned with CamScanner

Grouping,

$$\Rightarrow \langle A_i, (U+V)(U+V)^T\rangle^2 - \langle A_i, UU^T\rangle^2 - 2y_i\langle A_i, (U+V)(U+V)^T\rangle + 2y_i\langle A_i, UU^T\rangle$$

$$\Rightarrow \langle A_i, UU^T + UV^T + VU^T + VV^T\rangle^2 - \langle A_i, UU^T\rangle^2 - 2y_i\langle A_i, UU^T + UV^T + VU^T + VV^T\rangle + 2y_i\langle A_i, UU^T\rangle$$

$$\Rightarrow \left(\langle A_i, UU^T\rangle + \langle A_i, UV^T + VV^T + VU^T\rangle\right)^2 - \langle A_i, UU^T\rangle^2 - \cancel{2y_i\langle A_i, UU^T\rangle} - 2y_i\langle A_i, UV^T + VU^T + VV^T\rangle + \cancel{2y_i\langle A_i, UU^T\rangle}$$

$$\Rightarrow \cancel{\langle A_i, UU^T\rangle^2} + 2\langle A_i, UU^T\rangle\langle A_i, UV^T + VV^T + VU^T\rangle + \langle A_i, UV^T + VV^T + VU^T\rangle^2 - \cancel{\langle A_i, UU^T\rangle^2} - 2y_i\langle A_i, UV^T + VU^T + VV^T\rangle$$

$$\Rightarrow 2\langle A_i, UU^T\rangle\underline{\langle A_i, UV^T + VU^T + 0\rangle} + \underline{\langle A_i, UV^T + 0 + VU^T\rangle} - 2y_i\langle A_i, UV^T + VU^T + 0\rangle \quad [VV^T \text{ is neglegible}]$$

Similarly, multiplying UU^T with UV^T & VU^T is considerably small and can be neglected.

$$\therefore \Rightarrow 2\langle A_i, UU^T\rangle - 2y_i\langle A_i, UV^T + VU^T\rangle.$$

Taking trace of inner product.

$$\Rightarrow 2\langle A_i, UU^T\rangle - 2y_i\left(Tr[A_i^T UV^T] + Tr[A_i^T VU^T]\right)$$

$$\Rightarrow 2\langle A_i, UU^T\rangle - 2y_i\left(Tr[A_i^T UV^T] + Tr[U^T A_i^T V]\right)$$

Removing trace,

$$\Rightarrow 2\langle A_i, UU^T\rangle - 2y_i\left(\langle A_i U, V\rangle + \langle U A_i^T, V\rangle\right)$$

Neglecting V term again,

$$\Rightarrow 2\langle A_i, UU^T\rangle - 2y_i(A_i U + A_i^T U)$$

$$\Rightarrow 2\langle A_i, UU^T\rangle - 2y_i(A_i + A_i^T)U.$$

derivative (or) gradient of $f(U)$,

$$\boxed{\nabla f(U) = |2.\ \langle A_i, UU^T\rangle - 2y_i|(A_i + A_i^T)U}$$

Scanned with CamScanner

2c) $\min f(U) = \frac{1}{2m} \sum_{i=1}^{m} (\langle A_i, UU^T \rangle - y_i)^2$

$$L(U) = \frac{1}{2m} \sum_{i=1}^{m} (\langle A_i, UU^T \rangle - \langle A_i, UU^T \rangle)^2$$

we are given, rank of U is 1,

A is from random distribution with zero as mean and one as variance.

(m) no. of samples are infinite

Applying trace to $f(U)$,

$$f(U) = \frac{1}{2m} \sum_{i=1}^{m} (Tr(A_i, UU^T) - y)^2 \quad \text{— (1)}$$

$$y = Tr(A, X^*)$$

$$X^* = U^* U^{*T}$$

(1) can be written as,

$$E[f(U)] = \frac{1}{2}((Tr(A, UU^T)) - Tr(A, U^* U^{*T}))^2$$

$$\boxed{E[f(U)] = \frac{1}{2}(UU^T - U^* U^{*T})^2}$$

Adding V on both sides

$$f(U+V) = \frac{1}{2}(((U+V).(U+V^T)) - (U^*+V).(U^*+V^T))^2$$

From taylor expansion,

$$\nabla f(U) = 2(UU^T - U^* U^{*T}) U.$$

$$\Rightarrow 2(UU^T - U^* U^{*T}) U = 0. \text{ [setting gradient to 0]}$$

Since U is a rank-1 matrix, only 3 critical values for U,

$$U = 0, \quad U = \pm U^*$$

Taking Hessian of $f(U)$,

$$\nabla^2 f(U) = 4UU^T - 2U^* U^{*T} + 2\|U\|^2 I.$$

for $U = 0$,

$$\boxed{\nabla^2 f(U) = -2U^* U^{*T}}$$

$\nabla^2 f(U)$ is less than zero. It doesn't satisfy the condition for stationary point.

$\therefore \pm U^*$ are not critical points but only second order stationary points.

Scanned with CamScanner