２０２４年度　第１回　一橋大本番レベル模試

ソーシャルデータサイエンス・総合問題

解答・解説・採点基準

全３問　６０分　１００点満点

Ⅰ　（４０点）

【解答・採点基準】

問 1　$E_{\hat{a}} = \sum_{i=1}^{n} 2x_i\left(\hat{a}x_i + \hat{b} - y_i\right),$

$E_{\hat{b}} = \sum_{i=1}^{n} 2\left(\hat{b} + \hat{a}x_i - y_i\right)$

問 2　$\hat{a} = \dfrac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - \bar{x}^2},\ \hat{b} = \dfrac{\overline{x^2}\cdot\bar{y} - \bar{x}\cdot\overline{xy}}{\overline{x^2} - \bar{x}^2}$

問 3　データにノイズが乗っているとき，ノイズの影響を強く受けながら a_0～a_{n-1} を決定してしまうので，X の値から Y の値を正確に予測することができなくなる恐れがある。

問 4　λ の値を小さくしすぎた場合，問 3 と同様にノイズの影響を強く受けながら a_0～a_{n-1} を決定してしまうので，X の値から Y の値を正確に予測することができなくなる恐れがある。λ の値を大きくしすぎた場合，すでに取得した n 個のデータの影響に比べて，E'' の予測した陽関数の多項式の係数の影響が強くなり，すでに取得した n 個のデータをうまく利用す

問 1　8 点（各 4 点×2）

問 2　10 点（各 5 点×2）

問 3　8 点

問 4　14 点

ることができなくなる恐れがある。

問 3

＊ノイズの影響を受けながら係数を決定してしまうことが書かれていれば 5 点

＊値を正確に予測するのが困難になることが書かれていれば 3 点

問 4

＊λ の値を小さくしすぎた場合，問 3 と同じ結果となることが書かれていれば 6 点

＊λ の値を大きくしすぎた場合， 陽関数の多項式の係数の影響が強くなることが書かれていれば 5 点

＊既に取得したデータをうまく利用できないことが書かれていれば 3 点

2 （３０点）

【解答・採点基準】

問 1	8 回目
問 2	2 回目
問 3	線形探索：N 回目 二分法探索：$[\log_2 N]+1$ 回目
問 4	$\frac{1}{2}N(N-1)$ 回
問 5	優れている点：問 3 の計算結果から昇順または降順に並んでいる配列に対しては線形探索より比較回数が少なく処理にかかる時間が短い。 劣っている点：問 4 の計算結果から昇順または降順に並んでいない配列に対しては並び替えに多くの比較回数を必要としてしまい，線形探索より処理にかかる時間が長い。

問 1	4 点
問 2	4 点
問 3	7 点 ＊線形探索の場合を求めて 3 点 ＊二分法探索の場合を求めて 4 点
問 4	7 点
問 5	8 点

問 5

＊優れている点について， 昇順または降順に並んでいる配列に対しては比較回数が少ないことが書かれていれば 2 点。

＊優れている点について，処理にかかる時間が短いことが書かれていれば 2 点。

＊劣っている点について， 昇順または降順に並んでいない配列に対しては比較回数が多いことが書かれていれば 2 点。

＊劣っている点について，処理にかかる時間が長いことが書かれていれば 2 点。

3 （３０点）

【解答・採点基準】

問 1	c
問 2	低周波の基底画像のみを用いた場合にはおおむね元の画像が再現されるのに対し，高周波の基底画像のみを用いた場合には原形を留めないため，低周波の基底画像のみを用いて画像データを作成する手法を採用すべきである。
問 3	文字画像は高周波の基底画像を中心に構成されるので，文字の輪郭がぼやけてしまい，読みにくくなる。

問 1	10 点
問 2	10 点
問 3	10 点

問 2

＊低周波のみの画像の方が，高周波のみの画像よりも，元の画像に近い（画質が良い）という比較が書かれていれば 5 点。

＊低周波の基底画像のみを用いる方法を採用すべき旨が書かれていれば 5 点。

問 3

＊文字（の輪郭）が高周波の基底画像を中心に構成されることに言及していれば 5 点。

＊文字が読みにくくなることが述べられていれば 5 点。