1 Data Analysis

1. We are given an item space $X$ (the exact nature of the item is irrelevant for this problem). For an item $x \in X$ is also associated a numerical grade $y$, i.e., a real number. For simplicity, we assume $y \in Y$ where $Y$ is a finite subset of the real numbers. Let $P(x, y)$ be a probability distribution over $X \times Y$. For a fixed $n$, and a set of items $X_n = \{x_1, \ldots, x_n\}$ with the corresponding numerical grades $y_1, \ldots, y_n$, a ranking or an ordering of $X_n$ is simply a permutation $\sigma = (\sigma(1), \ldots, \sigma(n))$ of $(1, \ldots, n)$. For a given set of real numbers $c_1 > \ldots > c_n$, we use the following score to measure the quality of the ranking $\sigma$,

$$s(\sigma; \{(x_i, y_i)\}) = c_1 y_{\sigma(1)} + \ldots + c_n y_{\sigma(n)}.$$  

- Show a ranking $\sigma$ maximizes $s(\sigma)$ if

$$y_{\sigma(1)} > \ldots > y_{\sigma(n)},$$

i.e., we should order the items in $X_n$ by the decreasing order of the $y_i$’s.

- Now assume $(x_i, y_i), i = 1, \ldots, n$ are iid from $P(x, y)$, show that

$$p(y_i|x_1, \ldots, x_n) = p(y_i|x_i),$$

i.e., the conditional probability of $y_i$ given $x_1, \ldots, x_n$ is the same as the conditional probability of $y_i$ given $x_i$.

- A ranking function $R$ is a mapping from $X$ to the set of real numbers. We denote the ranking induced by $R$ as $\sigma_R$, where $\sigma_R$ is obtained by the decreasing order of $R(x_i), i = 1, \ldots, n$. For a ranking function $R$, define the expected score as

$$\mathcal{E}s(\sigma_R; \{(x_i, y_i)\})$$

where the expectation $\mathcal{E}$ is with respect to the product probability, i.e., $(x_i, y_i), i = 1, \ldots, n$ are iid from $P(x, y)$. Show that the following ranking function

$$R^*(x) = \sum_y yP(y|x)$$

maximizes the expected score.

2. From the above Problem, we know $R^*$ maximizes the expected score. We want to investigate whether the ranking (ordering) will change if we instead use

$$R_f^*(x) = \sum_y f(y)P(y|x)$$

as the ranking function, where $f$ is a strictly monotonically increasing function.
(a) Show if \( f \) is linear, using \( R^*(x) \) and \( R^*_j(x) \) as the ranking functions produces the same ranking (ordering).

(b) Show if \( y \) can take exactly two distinct values, for an arbitrary \( f \) which is strictly monotonically increasing, using \( R^*(x) \) and \( R^*_j(x) \) as the ranking functions produces the same ranking (ordering).

(c) Show (b) is not true if \( y \) can take more than two distinct values.