

レポート課題その1 説明

- 提出は、keio.jp で
- 次の項目を忘れないで下さい
 - 学籍番号、氏名
 - 「レポート課題1」というタイトル
 - 回答そのもの
 - 感想
- 各自、独自に行ってください。レポート作成も独自に行ってください。
- 締め切りは、11/27(月曜日)24時です。

レポート課題 1-1

- Laplace correction はなぜ行うのか、またその基本的な考え方を、サイコロ投げの課題(サイコロの各面が出る確率をその実験頻度から推定する課題)を例にとり、説明してください
- そして、実際にサイコロを投げてみて(といっても単にシミュレーション)、Laplace correction を行う必要があることを確かめてください(そして、報告してください)。
 - コイン投げを行う MsExcel ファイルを用意しました。これを少し直すだけでできます。10回試行を一組(これ一組で、サイコロのある面が出る確率を一回推定する)として100組(つまり100回推定する)使います。「一組」の試行数を変えたり、組数を変えたりしてみてください。
 - 「数式⇒計算手法の設定」を「手動」にしてあります。再計算を F9 で行い、ヒストグラムを作ります。グラフの再描画を行うには、棒グラフの棒の一つを左クリックし (fx欄に =SERIES(...) と表示されます)、fx欄にマウスカーソルを入れリターンを押下します。その状態のExcelファイルを保存して下さい。
 - そのデータを使って、一組当たりの試行数が少ないときには、特に、Laplace correction をした方が良い推定値が得られることを、実際の値で示してください。

レポート課題 1-2

- Naïve Bayes 法を用いて、下図左の訓練データが与えられたとき、「企画部門、30代前半、400万円台後半」および「販売部門、30代前半、600万円台後半」という未知サンプルの職位を推測して下さい。その答えが正しいような確率も教えてください。
 - なお、Rで行うと同時に、表を書いて、「手でも計算を行って下さい。

- この時、「項番」は属性に使ってはいけない。なぜか？
標準的なLaplace correctionを行ってください。

項番	部門	職位	年齢	給与
1	販売	2	30代前半	400万円台後半
2	販売	1	20代後半	200万円台後半
3	販売	1	30代前半	300万円台前半
4	設計	1	20代前半	400万円台後半
5	設計	2	30代前半	600万円台後半
6	設計	1	20代後半	400万円台後半
7	設計	2	40代前半	600万円台後半
8	企画	2	30代後半	400万円台後半
9	企画	1	30代前半	400万円台前半
10	庶務	2	40代後半	300万円台後半
11	庶務	1	20代後半	200万円台後半

データの数を数えて、推定

部門	職位	年齢	給与	回数
販売	1	20代後半	200万円台後半	1
販売	1	30代前半	300万円台前半	1
販売	2	30代前半	400万円台後半	1
設計	1	20代前半	400万円台後半	1
設計	2	30代前半	600万円台後半	1
設計	1	20代後半	400万円台後半	1
設計	2	40代前半	600万円台後半	1
企画	2	30代後半	400万円台後半	1
企画	1	30代前半	400万円台前半	1
庶務	2	40代後半	300万円台後半	1
庶務	1	20代後半	200万円台後半	1

一つの表に纏めておこう

部門	職位	年齢	給与	回数
販売	1	20代後半	200万円台後半	1
販売	1	30代前半	300万円台前半	1
販売	2	30代前半	400万円台後半	1
設計	1	20代前半	400万円台後半	1
設計	2	30代前半	600万円台後半	1
設計	1	20代後半	400万円台後半	1
設計	2	40代前半	600万円台後半	1
企画	2	30代後半	400万円台後半	1
企画	1	30代前半	400万円台前半	1
庶務	2	40代後半	300万円台後半	1
庶務	1	20代後半	200万円台後半	1

注！：Rを使うとき、スライドにあるプログラム中の「5」が「4」になるべきところがある