

# Машинное обучение в экономике

## Семинар 1. Байесовские сети

### Задание №1

У вас имеются данные, в которых хранится информация об уходе клиента, а также об уровне его образования и сокращении частоты использования приложения. Выборка и ее разбиение на части представлены в таблице.

Уход (Y)	1	1	1	0	0	1	1	0	0	0	1	1	0	1	0
Образование (O)	1	1	1	0	1	1	0	1	1	1	0	1	1	1	1
Сокращение (C)	1	0	1	1	0	0	1	1	0	0	1	0	0	1	0
Часть	1					2					3				

1. Вы используете Байесовский классификатор и прогнозируете, что клиент уйдет, если оценка условной вероятности данного события не меньше 0.55. Оцените точность байесовского классификатора с использованием двухчастной кросс-валидации.
2. Обучающая выборка была сформирована из первой и второй частей, а третья часть использовалась в качестве тестовой выборки. Порог прогнозирования остался таким же, как в предыдущем пункте. Оцените точность прогноза наивного байесовского классификатора на тестовой выборке, а также оцените условную вероятность и сделайте прогноз ухода клиента без образования, который не сократил частоту использования приложения. Аргументируйте, можно ли получить аналогичный прогноз с использованием обученного ранее байесовского классификатора.
3. Повторите предыдущий пункт, вместо наивного Байесовского классификатора используя Байесовскую сеть, в которой сокращения использования приложения является ребенком образования и родителем ухода клиента. Нарисуйте соответствующую Байесовскую сеть и сделайте вывод о целесообразности использования всех имеющихся признаков для прогнозирования ухода клиента.
4. В общем случае графически изобразите Байесовский классификатор и наивный Байесовский классификатор как Байесовскую сеть.

### Задание №2

События  $A$  и  $B$  независимы при условии события  $C$ .

1. Пусть  $P(A|C) = 0.3$ ,  $P(B, C) = 0.2$  и  $P(C) = 0.5$ . Найдите  $P(A, B|C)$ .
2. В дополнение к предыдущему пункту известно, что  $P(A, B) = 0.16$ . Рассчитайте  $P(C|A, B)$ .
3. В дополнение к предыдущим пунктам известно, что  $P(A|\bar{C}) = P(\bar{B}|\bar{C}) = 0.6$ . Проверьте, являются ли события  $A$  и  $B$  независимыми при условии события  $\bar{C}$ .

4. Вы решили использовать наивный Байесовский классификатор с двумя бинарными признаками  $X_{i1}$  и  $X_{i2}$  для прогнозирования бинарной целевой переменной  $Y_i$ . События  $X_{i1} = 1$ ,  $X_{i2} = 1$  и  $Y_i = 1$  эквивалентны событиям  $A$ ,  $B$  и  $C$  соответственно. Рассчитайте асимптотическое смещение используемой вами оценки условной вероятности  $P(Y_i = 1 | X_{i1} = 1, X_{i2} = 1)$ . Объясните причину возникновения этого смещения.
5. Для оценивания  $P(X_{i1} = 1, X_{i2} = 1, Y_i = 1)$  вы используете Байесовскую сеть, в которой  $X_{i1}$  является родителем  $X_{i2}$  и  $Y_i$ , которые, в свою очередь, не имеют детей. Найдите асимптотическое смещение этой оценки.
6. Докажите следующие равенства:

$$P(A|C) = P(A|C, B) \quad P(B|C) = P(B|C, A)$$