

4-1-1 Исправить название фрагмента на «Определение мультиколлинеарности» (сейчас определение)

http://youtu.be/9MhZ0UCb_Cc

0:08 убрать слова «кафедра публичной политики»

0:52 заменить появляющуюся матрицу на более длинную:

$$\begin{pmatrix} 1 & 4 & 12 & 8 \\ 1 & 3 & 3 & 3 \\ 1 & 1 & 7 & 4 \\ 1 & 2 & 4 & 3 \\ 1 & 3 & 5 & 4 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

3:53 убрать полностью «Практика» и два подпункта

6:07 ошибка в начале формулы (пропущен квадрат), должно быть:

$$se^2(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{RSS_j} = \frac{\hat{\sigma}^2}{TSS_j \cdot (1-R_j^2)} = \frac{1}{1-R_j^2} \frac{\hat{\sigma}^2}{TSS_j}$$

9:57 ошибка в начале формулы (пропущен квадрат), должно быть:

$$se^2(\hat{\beta}_j) = VIF_j \cdot \frac{\hat{\sigma}^2}{TSS_j}$$

10:19 после пункта «Выборочные корреляции...» добавить формулу

$$sCorr(x, z) = \frac{\sum (x_i - \bar{x})(z_i - \bar{z})}{\sqrt{sVar(x) \cdot sVar(z)}}$$

10:32 ничего нового не появляется

10:36 старый текст очищается, появляется новый (это исправленный вариант того, что было в 10:32):

Некоторые источники считают признаком мультиколлинеарности:

* $VIF_j > 10$

* $sCorr(x, z) > 0.9$

10:57 фрагмент видео обрывается неожиданно на незаконченной фразе

4-1-2 Что поделаться с мультиколлинеарностью?

<http://youtu.be/yXQGrWGVpoo>

0:08 убрать слова «кафедра публичной политики»

1:21 Исправить второй пункт на «Уменьшить дисперсию оценок, пожертвовав их несмещенностью»

1:28 Исправить третий пункт на «Мечта: уменьшить дисперсию оценок, используя больше наблюдений»

2:05 изменяем заголовок слайда на «Жертвуем несмещенностью, чтобы снизить дисперсию»

2:10 надпись «Модель зависит от всех регрессоров» убираем полностью

2:15 Исправить надпись на

* Выкинуть часть регрессоров.

Жертвуем: знанием выкидываемых коэффициентов, несмещенностью оставшихся коэффициентов.

2:23 Исправить надпись на

* Использовать МНК со штрафом

Жертвуем: несмещенностью коэффициентов

2:24 повторную надпись «Жертвуем несмещенностью!» убираем полностью

4-1-3 изменить название фрагмента на «Ридж и LASSO регрессия»

<http://youtu.be/fwqs8mPEZFc>

0:08 убрать слова «кафедра публичной политики»

0:23 появляется надпись:

Общий принцип:

$$\min_{\hat{\beta}} RSS + \text{Штраф}$$

0:46 про общий принцип надпись стираем, теперь появляется:

Ридж-регрессия (ridge)

$$\min_{\hat{\beta}} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda(\hat{\beta}_1^2 + \hat{\beta}_2^2 + \dots + \hat{\beta}_k^2)$$

0:57 дополнительно появляется:

LASSO регрессия

$$\min_{\hat{\beta}} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda(|\hat{\beta}_1| + |\hat{\beta}_2| + \dots + |\hat{\beta}_k|)$$

1:10 дополнительно появляется:

Метод эластичной сети

$$\min_{\hat{\beta}} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda_1 \sum_{j=1}^k |\hat{\beta}_j| + \lambda_2 \sum_{j=1}^k \hat{\beta}_j^2$$

4-1-4 Идея метода главных компонент

<http://youtu.be/6LnUUI0Iotw>

0:08 убрать слова «кафедра публичной политики»

2:29 исправить на:

* переменная pc_1 имеет максимальную выборочную дисперсию $sVar(pc_1)$

3:04 исправить на:

* переменная pc_2 некоррелирована с pc_1 и имеет максимальную $sVar(pc_2)$

3:33 исправить на:

* переменная pc_3 некоррелирована с pc_1 , pc_2 и имеет максимальную $sVar(pc_3)$

* и т.д.

4:30 выше фразы «Первая главная компонента — математика» добавить слово «Упрощённо:»

5:07 — 5:14 процесс рисования облака точек можно ускорить

4-1-5 Пример нахождения главной компоненты

<http://youtu.be/oRYuXjZJip0>

0:08 убрать слова «кафедра публичной политики»

4-1-6 Свойства главных компонент

<http://youtu.be/-UcSI47ITfU>

0:08 убрать слова «кафедра публичной политики»

0:35 исправить вторую формулу (вместо нижнего индекса k должен быть 2), в результате должно получиться:

$$pc_1 = v_{11} \cdot x_1 + v_{21} \cdot x_2 + \dots + v_{k1} \cdot x_k$$

$$pc_2 = v_{12} \cdot x_1 + v_{22} \cdot x_2 + \dots + v_{k2} \cdot x_k$$

...

0:42 исправить формулу на

$$sCorr(pc_j, pc_m) = 0$$

0:52 исправить формулу на

$$sVar(x_1) + sVar(x_2) + \dots + sVar(x_k) = sVar(pc_1) + sVar(pc_2) + \dots + sVar(pc_k)$$

2:13 — 2:43 — удалить полностью (эти же слова более аккуратно я говорю с момента 2:43). В надпись, появляющуюся в 2:13 добавить тире, чтобы вышло:

Если: все переменные центрированы, $\bar{x}_j = 0$

То: $pc_j = X \cdot v_j$ и $|pc_j|^2 = \lambda_j$,

где λ_j — собственные числа,

v_j — собственные вектора матрицы $X'X$

5:40 исправить второй пункт на

* Бездумное применение перед регрессией

6:13 исправить знаменатель в формуле в конце текста, должно быть:

$$x_j = \frac{a_j - \bar{a}_j}{se(a_j)}$$

7:10 появляется подзаголовок

Применение метода главных компонент перед регрессией

7:21 ниже подзаголовка появляется надпись:

Небезопасная процедура:

Шаг 1. Найти главные компоненты, pc_1, pc_2

7:28 вместо текущего текста появляется

Шаг 2. Построить регрессию y на pc_1, pc_2

7:39 вместо текущего текста появляется:

Проблемы:

* коэффициенты при pc_j сложнее интерпретировать

8:03 добавляется строка

* самый изменчивый регрессор x гипотетически может быть наименее связан с y

8:43 Вместо «Мораль — мультиколлинеарность» оставляем просто «Мультиколлинеарность»

8:46 исправляем пункт на

* сильная линейная зависимость между регрессорами

4-2-1 Доверительные интервалы при мультиколлинеарности

<http://youtu.be/9P4pCJN9ENw>

0:08 убрать слова «кафедра публичной политики»

переснять, т.к. нет `lab_04_before.R`

1:22 показать график

4-2-2 Lasso регрессия

<http://youtu.be/ixdKGZJeSt0>

0:08 убрать слова «кафедра публичной политики»

4-2-3 Ридж-регрессия и идея оценки лямбды

<http://youtu.be/dAJu-BUyvks>

0:08 убрать слова «кафедра публичной политики»

4-2-4 Метод главных компонент

<http://youtu.be/g83D9xx0WzE>

0:08 убрать слова «кафедра публичной политики»