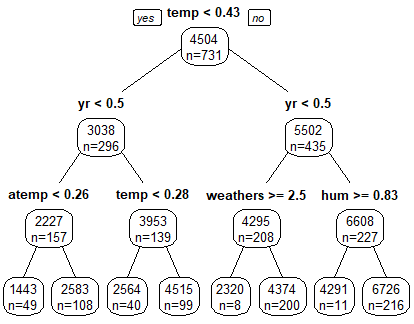
**Примерен Тест 2**

**Задача 1: Във връзка с оценяването на броя хора на мероприятие на открито в дадено градче е получено следното дърво:**



**temp е очакваната температурата през деня; atemp е средната температура през този ден на годишна база; yr<0.5 означава сезони зима, пролет, а иначе лято,есен; hum е влажността; weathers е осветеността (по-високи стойности, по-малко облаци).**

**А) Колко листа има дървото?**

**Б) Колко е максималната дълбочина на дървото?**

**В) Кой е най-важният фактор отговорен за посещаемостта на дадено събитие на открито?**

**Г) Влияят ли сезоните на броя хора?**

**Д) Какво означава стойността оградена във всяко листо на дървото и символа “n=” под нея?**

**Е) Ако е очакваната температура за дена(temp) е под 0.28 и е лято, колко хора се очакват, ако оргнизираме събитие на открито?**

**Отговори:**

А) 8 листа;

Б) 3 защото толкова е максималния брой правила преди да стигнем до финален резултат;

В) Нaй-важният фактор е в „корена“ на дървото, т.е. в началото и това е temp.

Г) Сезонте са част от дървото и влияят съществено;

Д) Оградената стойност означава средната стойност на извадката определена от листото, а стойността под нея е броя на елементите в тази извадка;

Е) 4515 човека

**Задача 2: Имаме следните грешки от Модел 1 и Модел 2:**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Модел 1:** | 1.5 | 2 | 3 | -1 | 5 | 4 | 1 | 8 | -3 | 0.5 |
| **Модел 2:** | 1.8 | 1 | -0.2 | -2 | -3 | -2 | 6 | -5 | 0 | -7 |

**Определете кой модел е по-добър на база на средната абсолютна грешка.**

**Отговор:**

Модел1: (1.5+2+3+1+5+4+1+8+3+0.5)/10= 2.9

Модел2: (1.8+1+0.2+2+3+2+6+5+0+7)/10=2.8

Модел 2 е с по-ниска грешка и следователно е предпочитан.

**Задача 3: Ако изглаждаме с пълзащо средно чрез две стойности назад, изгладете следните данни:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Времеви ред:** | **1** | **1.3** | **1.6** | **1.4** | **2** |
| **Изгладен:** |  |  |  |  |  |

**Отговор:**

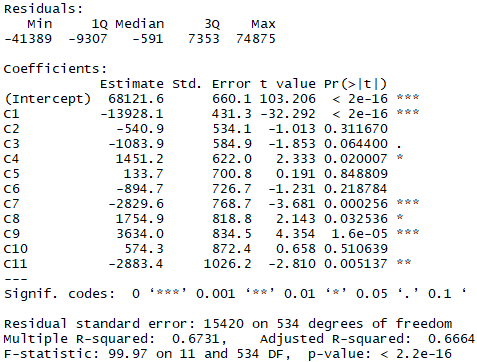
Понеже изпозлваме последните две стойности преди изгладената, не може да изгладим началните две стойнсоти на реда и продължаваме от третата:

V(3) = (V(1)+V(2)+V(3))/3 = (1+1.3+1.6)/3=3.9/3 = **1.3**

V(4) = (V(2)+V(3)+V(4))/3 =(1.3+1.6+1.4)/3=4.3/3 = **1.433**

V(5) = (V(3)+V(4)+V(5))/3 = (1.6+1.4+2)/3= 5/3 = **1.66**

**Задача 4: Нека предположим, че сме приложили метода на главните компоненти върху моделните променливи на даден регресионен модел и след това сме заместили пръвоначалните променливи с тях:**

****

**А) Можете ли да познаете колко са били моделните променливив като брой по броя на компонентите?**

**Б) Можете ли да познаете колко е била модлената точност R квадрат преди трансформацията?**

**В) Дава ли регресията признаци, че е възможно редуциране на размерността и защо?**

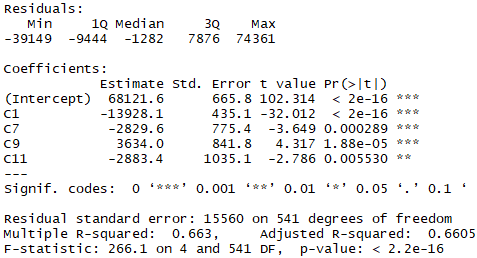
**Отговор:**

А) Броят на моделните променливи съвпада с броя на компонентите, т.е. 11.

Б) При използване на всички компоненти регресионния модел остава Непроменен.

В) Много компоненти са лошо оцененни, което означава, че могат да се премахнат и най-вероятно това няма да полияе на регресията, т.е., да, може да се псотиген значителна редукция.

**Задача 5: След отстраняване на някои от компонентите от предходната задача имаме:**

****

**А) Значителна ли е редукцията на размерността?**

**Б) Какво е качеството на получения модел спрямо предходния?**

**Отговор:**

А) Да, от 11 променливи имаме 4, което е близо 3 пъти редукция на размерността. Всяка редукция над 5% е значима.

Б) Моделите са съпоставими по модлени статистики. Разликите са много малки, а вторият модел е значитено по-опростен.

**Задача 6: Имaме следните резултати от даден модел:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Реални стойности** | **9** | **6** | **5** | **10** | **4** | **8** | **3** |
| **Моделни стойности** | **8** | **7** | **5** | **7** | **2** | **4** | **6** |

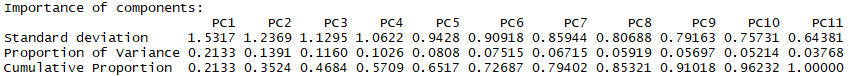
**Намерете колко е средната процентна грешка (MAPE) на модела.**

**Отговор:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Грешка** | **1** | **-1** | **0** | **3** | **2** | **4** | **-3** |
| **% грешка** | **1/9**  **= 11.11%** | **-1/6**  **= 16.66%** | **0**  **= 0%** | **3/10**  **= 30%** | **2/4**  **= 50%** | **4/8**  **= 50%** | **-3/3**  **= 100%** |

Средната грешка е: (11.11%+16.66%+0%+30%+50%+50%+100%)/7= **36.82%.**

**Задача 7: Разполагате със следния изход от метода на главните компоненти:**



**А) Колко процента от данните обяснява 6 –тата компонента?**

**Б) Колко процента от данните обясняват първите 8 компоненти?**

**В) Ако искаме да си запазим информативността на данните до поне 90%, колко компоненти трябва да оставим?**

**Отговор:**

А) Приблизително 7.52%.

Б) Около 85%.

В) Виждаме, че първите 9 компоненти ни дават 91%, т.е. поне 9.

**Задача 8: Знаем, че интерпретирането на даден регресионен модел е много важен елемент от анализа му.**

**А) По какво преценяваме колко влиятелна е дадена моделна променлива?**

**Б) Може ли да се подведем при интерпретациите от А) ?**

**В) Как да избегнем това проблема от Б)?**

**Отговор:**

А) Колко е влиятелна дадена променлива зависи от абсолютната стойност на коефициента пред нея, но внимавайте със скалите.

Б) При различни скали на данните или при значими разлики в големината на стойностите на променливите.

В) Следва да се приложи Мин-Макс нормализация (или поне по стандартно отклонение) на входните променливив преди да се премине към интерпретиране.

**Задача 9. Имаме следните данни за БВП:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Година** | **БВП** | **Година** | **БВП** | **Година** | **БВП** |
| **1990** | 52908 | **1994** | 45057 | **1998** | 48199 |
| **1991** | 48440 | **1995** | 46345 | **1999** | 45495 |
| **1992** | 44917 | **1996** | 47087 | **2000** | 27137 |
| **1993** | 44252 | **1997** | 46569 | **2001** | 29236 |

1. **Изчислете индекс на БВП с база 1995 година.**
2. **Колко се е изменило БВП 1997 спрямо 1995.**
3. **Има ли подобрение на икономиката от 1995 до 2001?**

Решение:

1. Всяка стойност делим на базовата стойност от 1995 и умножаваме по 100.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Година** | **БВП индекс** | **Година** | **БВП индекс** | **Година** | **БВП индекс** |
| **1990** | 114.16 | **1994** | 97.22 | **1998** | 104.00 |
| **1991** | 104.52 | **1995** | 100.00 | **1999** | 98.16 |
| **1992** | 96.92 | **1996** | 101.60 | **2000** | 58.55 |
| **1993** | 95.48 | **1997** | 100.48 | **2001** | 63.08 |

б) Индекс 1997 = 100.48, а 1995=100, измемението е +0.48%.

в) През 2001 имаме индекс 63.08 спрямо базата ни, т.е. БВП-то ни е 63.08% от първоначалното, далеч не може да говорим за подобрение.

**Задача 10. Разполагаме с данни за следния времеви ред:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Период:** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **Стойност:** | **1** | **1.3** | **1.2** | **1.6** | **1.7** | **1.8** | **1.4** |

**а) Изгладете данните с пълзяща средна МА с времеви прозорец 4.**

**б) Модифицираме експоненциалната пълзяща средна ( тази е за прогнозиране, а на слайдовете съм ви дал за изглаждане) до следната формула:**

**Така модифицираната формула вече може да се използва за прогнозиране един интервал напред във времето. Ако , то прогнозирайте Период 5.**

**В) А какво ще стане с Период 5, ако параметрите на модела са ? Колко е грешката на модела за Период 5?**

**Решение:**

**а)**

**б)** Имаме следната формула:

Искаме да намерим

, следователно:

.

в)