《生物医学信息》课程练习

第一题：生存分析

假设初始状态有12个个体，事件发生时间如下：

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Time | 5 | 10 | 15 | 20 | 25 | 35 | 35 | 45 | 55 | 65 | 75 |
| Event | F | F | C | F | C | F | F | F | C | F | F |

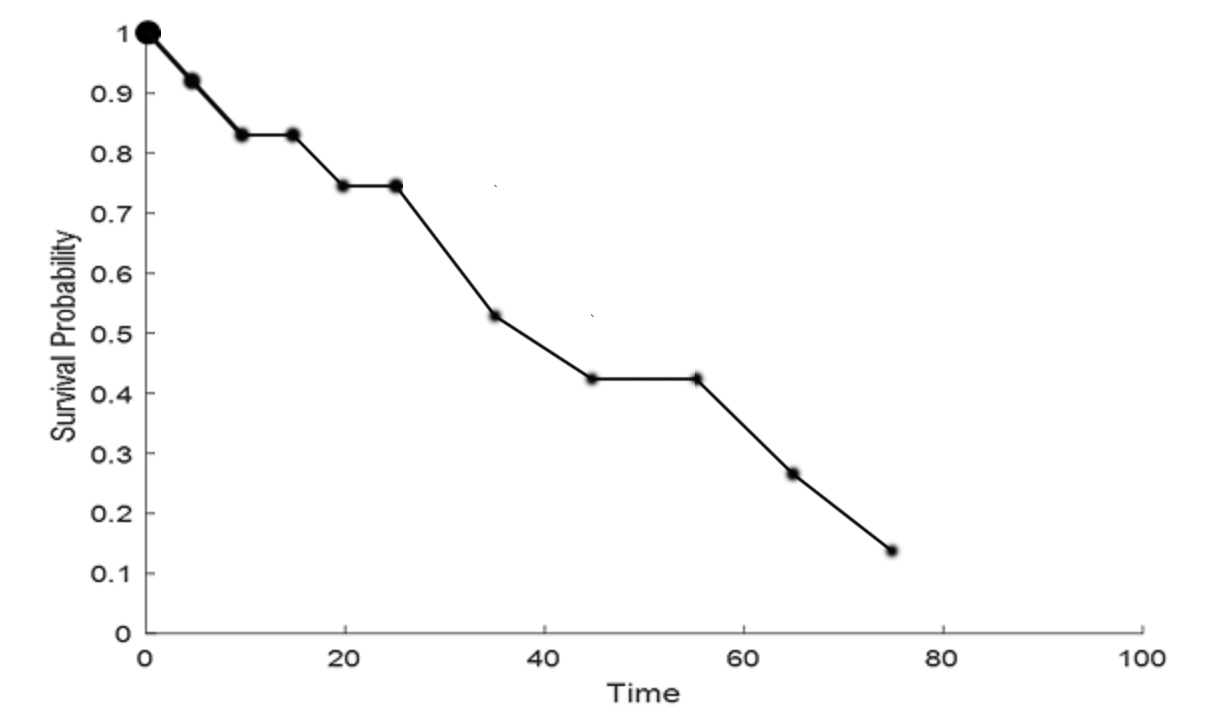
F: Failure， C： Censored

请根据公式: 

计算如下生存分析表，并绘制生存曲线。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Time | #at risk | #failure | #censored | S(t) |
| 0 | 12 | 0 | 0 | 1 |
| 5 | 12 | 1 | 0 | 11/12 |
| 10 | 11 | 1 | 0 | 5/6 |
| 15 | 10 | 0 | 1 | 5/6 |
| 20 | 9 | 1 | 0 | 20/27 |
| 25 | 8 | 0 | 1 | 20/27 |
| 35 | 7 | 2 | 0 | 100/189 |
| 45 | 5 | 1 | 0 | 80/189 |
| 55 | 4 | 0 | 1 | 80/189 |
| 65 | 3 | 1 | 0 | 160/567 |
| 75 | 2 | 1 | 0 | 80/567 |
|  |  |  |  |  |

提示：丢失事件不改变S(t)但改变分母。两种事件在图中各用一种标记标出。



第二题：朴素贝叶斯分类

已知以下10个乳腺肿瘤数据样本：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **X1, 肿块厚度** | **X2, 细胞核** | **X3, 有丝分裂** | **类别** |
| 1 | 重度 | 严重 | 正常 | 恶性 |
| 2 | 轻度 | 正常 | 正常 | 良性 |
| 3 | 轻度 | 正常 | 重度 | 良性 |
| 4 | 轻度 | 严重 | 重度 | 恶性 |
| 5 | 轻度 | 正常 | 正常 | 良性 |
| 6 | 重度 | 严重 | 中度 | 恶性 |
| 7 | 重度 | 正常 | 中度 | 恶性 |
| 8 | 重度 | 正常 | 正常 | 良性 |
| 9 | 轻度 | 正常 | 中度 | 良性 |
| 10 | 重度 | 严重 | 重度 | 恶性 |

请按如下步骤建立朴素贝叶斯分类器识别良性C1和恶性肿瘤C2。 可自行对变量取值编码，填在上表中。

（1）计算每个类的先验概率P(Ci)，其中，C1为良性， C2为恶性。

|  |  |
| --- | --- |
| 概率 | 概率值 |
| P(C1) | 1/2 |
| P(C2) | 1/2 |

（2）计算条件概率（请自行补齐空缺行）：

|  |  |
| --- | --- |
| 概率 | 概率值 |
| P(X1=重度|C1) | 1/5 |
| P(X1=重度|C2) | 4/5 |
| P(X1=轻度|C1) | 4/5 |
| P(X1=轻度|C2) | 1/5 |
| P(X2=严重|C1) | 0 |
| P(X2=严重|C2) | 4/5 |
| P(X2=正常|C1) | 1 |
| P(X2=正常|C2) | 1/5 |
| P(X3=正常|C1) | 3/5 |
| P(X3=正常|C2) | 1/5 |
| P(X3=中度|C1) | 1/5 |
| P(X3=中度|C2) | 2/5 |
| P(X3=重度|C1) | 1/5 |
| P(X3=重度|C2) | 2/5 |

(3) 分别对数据X1 = {肿块厚度=轻度, 细胞核=严重, 有丝分裂=中度}和X2={肿块厚度=轻度, 细胞核=正常, 有丝分裂=正常}，列出分类概率P(C1|X)和P(C2|X)的计算公式及过程，并计算最终分类结果。

P(C1|X1)=P(X1|C1)\*P(C1)/P(X1)

=P(肿块厚度=轻度|C1)\*P(细胞核=严重|C1)\*P(有丝分裂=中度|C1)\*P(C1)/P(肿块厚度=轻度)/P(细胞核=严重)/P(有丝分裂=中度)

=0.8\*0\*0.2\*0.5/0.5/0.4/0.3

=0

P(C2|X1)=P(X1|C2)\*P(C2)/P(X1)

=P(肿块厚度=轻度|C2)\*P(细胞核=严重|C2)\*P(有丝分裂=中度|C2)\*P(C2)/P(肿块厚度=轻度)/P(细胞核=严重)/P(有丝分裂=中度)

=0.2\*0.8\*0.4\*0.5/0.5/0.4/0.3

=0.53

P(C1|X2)=P(X2|C1)\*P(C1)/P(X2)

= P(肿块厚度=轻度|C1)\*P(细胞核=正常|C1)\*P(有丝分裂=正常|C1)\*P(C1)/P(肿块厚度=轻度)/P(细胞核=正常)/P(有丝分裂=正常)

=0.8\*1\*0.6\*0.5/0.5/0.6/0.4

=0.2

P(C2|X2)=P(X2|C2)\*P(C2)/P(X2)

= P(肿块厚度=轻度|C2)\*P(细胞核=正常|C2)\*P(有丝分裂=正常|C2)\*P(C2)/P(肿块厚度=轻度)/P(细胞核=正常)/P(有丝分裂=正常)

=0.2\*0.2\*0.2\*0.5/0.5/0.6/0.4

=0.033