

CE225 - Modelos Lineares Generalizados

Cesar Augusto Taconeli

23 de outubro, 2018

Aula 12 - Regressão para dados binários - predição

- Modelos de regressão para dados binários são bastante utilizados para predição, ou seja, classificar indivíduos conforme suas probabilidades estimadas.
- Alguns exemplos:
 - Predição (classificação) de clientes em bons ou maus pagadores;
 - Predição de e-mails em spams ou não spams;
 - Predição do resultado de um jogo de basquete (vitória do time mandante ou do time visitante);
 - Prognóstico de um paciente (cura ou não cura)...

- É fortemente recomendável avaliar o poder preditivo do modelo ajustado com dados que não foram usados no ajuste.
- Ajustar o modelo e avaliar a qualidade preditiva usando os mesmos dados tende a produzir resultados excessivamente otimistas.
- Algumas possibilidades:
 - Separar aleatoriamente a amostra em duas partes (uma para ajuste, a outra para predição);
 - Usar validação cruzada (caso particular: *leave one out*).

Validação cruzada

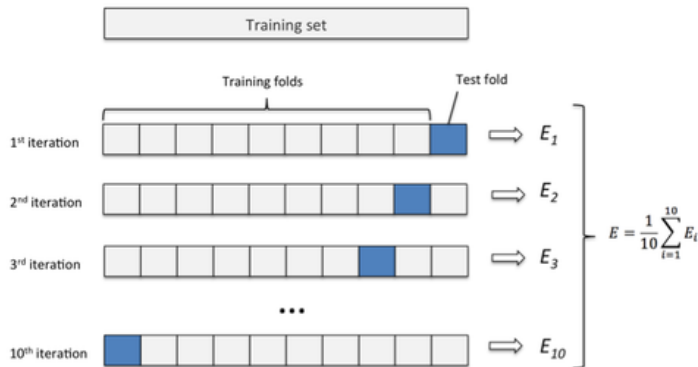


Figura 1: Ilustração - validação cruzada

- Sejam $\hat{\pi}_i$ as estimativas de $P(y_i = 1)$, $i = 1, 2, \dots, n$.
- Considere uma regra do tipo:
 - Predizer $\hat{y}_i = 0$ se $\hat{\pi}_i \leq p_0$;
 - Predizer $\hat{y}_i = 1$ se $\hat{\pi}_i > p_0$,

para algum valor (ponto de corte) especificado p_0 e $i = 1, 2, \dots, n$.

- É comum (mas não obrigatório) usar $p_0 = 0.5$, classificando pelo resultado com maior probabilidade.
- Diferentes valores de p_0 conduzem a diferentes regras de predição.

Tabelas de classificação

- Dadas as predições e os valores realmente observados de y , podemos construir uma tabela de classificação.

Tabela 1: Tabela de classificação

y	\hat{y}	
	0	1
0	n_{00}	n_{01}
1	n_{10}	n_{11}

Sumarizando o poder preditivo

- A **acurácia** de um modelo (ou de uma regra de classificação) é definida pela probabilidade de classificação correta, isto é,

$$Acur = P(\hat{y} = 0, y = 0) + P(\hat{y} = 1, y = 1);$$

- A **sensibilidade** de um modelo é definida pela probabilidade de classificar como *sucesso* dado que se trata, de fato, de um *sucesso*:

$$Sens = P(\hat{y} = 1|y = 1);$$

- A **especificidade** de um modelo é definida pela probabilidade de classificar como *fracasso* dado que se trata, de fato, de um *fracasso*:

$$Esp = P(\hat{y} = 0|y = 0).$$

Sumarizando o poder preditivo

- Podemos estimar acurácia, sensibilidade e especificidade com base nas frequências de uma tabela de classificação:

$$\widehat{Acur} = \frac{n_{00} + n_{11}}{n_{00} + n_{01} + n_{10} + n_{11}};$$

$$\widehat{Sens} = \frac{n_{11}}{n_{10} + n_{11}};$$

$$\widehat{Esp} = \frac{n_{00}}{n_{00} + n_{01}}.$$

Sumarizando o poder preditivo

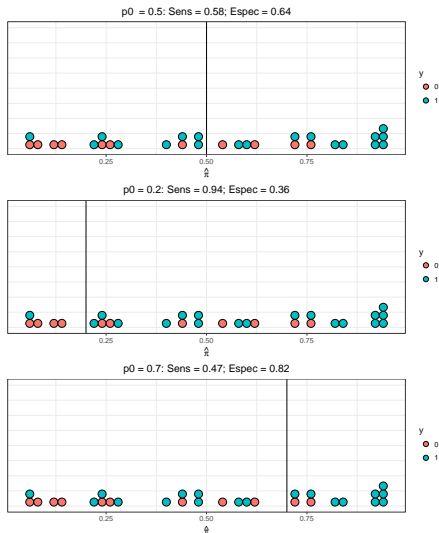


Figura 2: Ilustração - predição para dados binários

Curva ROC

- Uma forma de analisar o poder preditivo associado a diferentes regras de decisão (valores de p_0) é por meio da **curva ROC**.
- A curva ROC permite avaliar conjuntamente a sensibilidade e a especificidade para diferentes valores de p_0 .
- Para valores $p_0 \approx 1$, temos sensibilidade próxima de zero e especificidade próxima de um;
- Para valores $p_0 \approx 0$, temos sensibilidade próxima de um e especificidade próxima de zero;
- Em geral, busca-se p_0 tal que se tenha, conjuntamente, elevadas sensibilidade e especificidade;
- A **área sob a curva ROC** é uma medida de poder preditivo do modelo.

Curva ROC

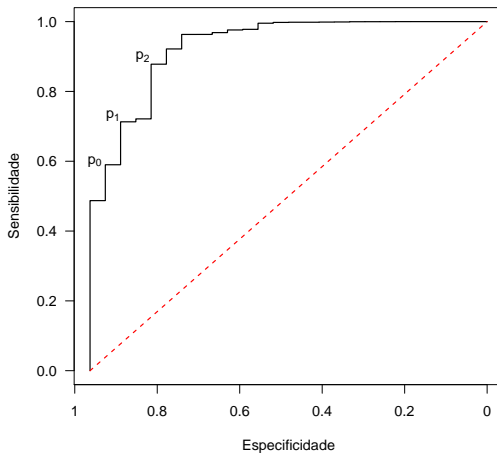


Figura 3: Ilustração - Curva ROC

Escolha da regra de decisão

- **Regra de Youden:** O ponto de corte ótimo é aquele mais distante da distância da reta identidade da curva ROC. Corresponde a p_0 tal que:

$$Sens + Esp$$

é máximo.

- **Closest topleft:** Neste caso, o ponto de corte ótimo é o mais próximo do canto superior esquerdo da curva ROC. Corresponde a p_0 tal que:

$$(1 - Sens)^2 + (1 - Esp)^2$$

é mínimo.

Escolha da regra de decisão - incorporando custos

- Ambos os critérios podem ser modificados de forma a incorporar custos de classificações incorretas e a prevalência de “sucessos” na população.

- Sejam:

- 1 c : o custo relativo de um falso negativo comparado ao de um falso positivo:

$$c = \frac{\text{custo}(-|+)}{\text{custo}(+|-)};$$

- 2 π : a prevalência (ou proporção) de sucessos (+) na população.

Escolha da regra de decisão - incorporando custos

- **Regra de Youden:** O ponto de corte ótimo é aquele que maximiza:

$$Sens + r \times Esp.$$

- **Closest topleft:** Neste caso, o ponto de corte ótimo minimiza:

$$(1 - Sens)^2 + r \times (1 - Esp)^2,$$

em que:

$$r = \frac{1 - \pi}{c\pi}.$$