بسم الله الرحمن الرحیم

Model Selection Information Criteria
معیارهای انتخاب مدل

ارائه دهندگان: آیدا تاج‌دینی، سیمین کردی

درس مربوطه: مدل خطی

دی ۹۷

فهرست مطالب

تعریف
هدف و اهمیت انتخاب مدل
معیارهای انتخاب مدل
مراجع

اهداف تجزیه و تحلیل داده ها

- استخراج اطلاعات از داده ها
- پیش بینی پاسخ ها با استفاده از متغیرهای پیش بینی‌گر

مطالعات

آزمایش‌های کنترل شده

- کنترل متغیرهای توضیحی به طور کامل صورت می گیرد
- بین متغیر های پیش بینی گر همبستگی وجود نخواهد داشت(یک وضعیت ایده آل)

گروه مورد آزمایش

گروه کنترل

مشاهداتی

- بین متغیرها همبستگی چندگانه وجود دارد
- همه متغیر ها مشاهده می شوند و هیچ کنترلی صورت نمیگیرد
- بیشترین نیاز را به Model Selection دارند

❑ سوال های مطرح شده در انتخاب مدل :

۱. چه تعداد متغیر لازم است؟ ← زیر مجموعه ای از متغیر ها در نظر می گیریم که واریانس و پراکندگی کمتری دارند.

۲. با توجه به حجم زیرمجموعه ها چه متغیر هایی باید انتخاب شود؟ ← به کمک یک سری رویکرد ها و آماره ها برای مقایسه(این رویکرد و آماره ها را در معیار های انتخاب مدل توضیح می دهیم).

❑ مثال هایی از مدل های مختلف:

- مدل های خطی
- مدل های آرما(ARMA) برای داده های سری زمانی
- توزیع وایبل برای داده‌های Time-to-event
- رگرسیون با خطاهای خودهمبسته

❑ به‌طور کلی، مدل‌سازی آماری شامل:

برآورد مدل | انتخاب مدل | اعتبار سنجی مدل | ارزیابی و بررسی مدل

هدف از مدل سازی

- دریافت قوانین کلی از داده ها
- داده یک نمونه از جامعه است

بهترین مدل

- عملکرد پیش بینی بالایی دارد
- متغیرهای پیش بینی کننده مهم را تشخیص می دهد
- اثر این پیش بینی کننده ها را بررسی میکند

یک مدل خوب

- استفاده از آن آسان و قابل فهم باشد

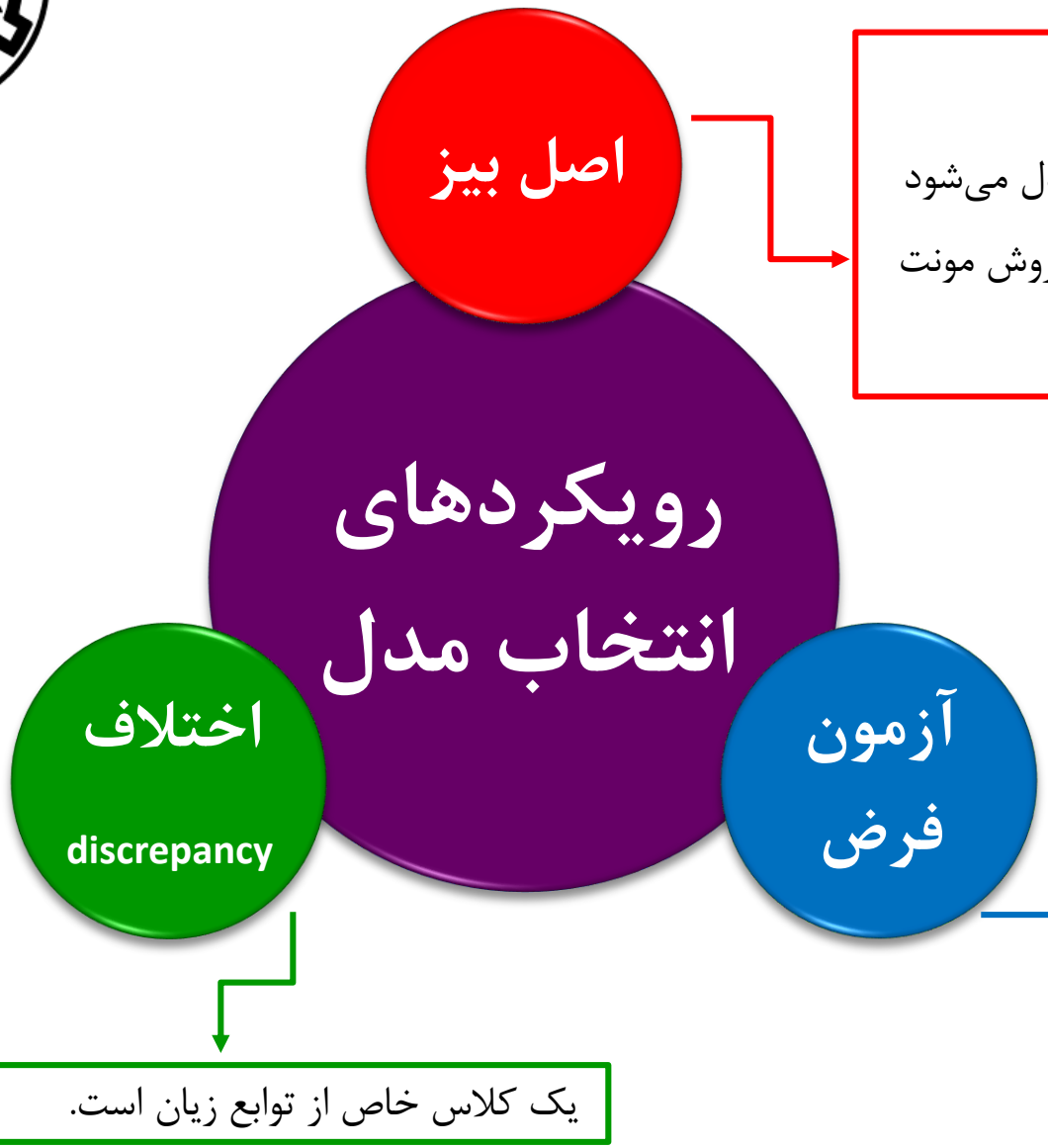
مراحل انتخاب مدل

الگوریتم

برای تولید مدل‌های مختلف به‌عنوان مجموعه ای از مدل های کاندید شده که شامل بهترین زیر مجموعه هاست ایجاد می‌کنیم

معیار

انتخاب و ارزیابی مدل ها را بررسی می‌کنیم



دو نوع روش ارائه می دهد:

۱- معیار اطلاعات بیزی BIC است، که توسط تقریب احتمال پسین مدل می‌شود

۲- انتخاب مدل بیزی که براساس محاسبه احتمال پسین با استفاده از روش مونت کارلو است

یک کلاس خاص از توابع زیان است.

معمولا فرض می‌کند که مدل‌های کاندید شده درست و شامل یک مدل واقعی می‌شوند.

- **معیارهایی که نیکویی برازش را کنترل می‌کنند**

R^2_{adj}

- $R^2_{adj} = 1 - \frac{n-1}{n-p}\left(\frac{SSE}{SSTO}\right)$ & $R^2_{adj} = 1 - \frac{\frac{MSE}{SSTO}}{n-1} = 1 - \frac{MSE}{Constant}$

R^2_p

- $R^2_p = 1 - \frac{SSE_p}{SST} = \frac{SSR}{SSTO}$

MSE

- $MSE = \frac{SSE}{n-p}$
- از بین چند مدل کاندید، مدلی که MSE کمتر و R^2_p و R^2_{adj} بیشتری داشته باشد، بهتر است.

- **معیارهایی که نیکویی برازش را کنترل می‌کنند**

VIF

$$X=\begin{pmatrix} x_{11} & x_{12} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{pmatrix} \qquad X^*=\begin{pmatrix} \frac{x_{11}-\bar{x_1}}{s_1} & \frac{x_{12}-\bar{x_2}}{s_2} \\ \vdots & \vdots \\ \frac{x_{n1}-\bar{x_1}}{s_1} & \frac{x_{n2}-\bar{x_2}}{s_2} \end{pmatrix}$$

$$X^*X'^*=\begin{pmatrix} 1 & r_{12} \\ r_{21} & 1 \end{pmatrix} \qquad (X^*X'^*)^{-1}=\begin{pmatrix} \frac{1}{1-r_{12}^2} & \frac{-r_{12}}{1-r_{12}^2} \\ \frac{-r_{12}}{1-r_{12}^2} & \frac{1}{1-r_{12}^2} \end{pmatrix}$$

$$VIF_i = \frac{1}{1-R_i^2}$$

VIF ایده آل برابر ۱ است، زمانی که r برابر صفر باشد.

- معیارهایی که کیفیت پیش بینی مدل را ارزیابی می‌کنند

$Mallow's\ C_p$:

- $C(p) = \frac{SSE_p}{MSE(full)} - (n - 2p)$

A model is considered “good” if → $C(p) \leq p$

$PRESS$:

- $PRESS = \sum_{i=1}^{n} e_i^2 - i = \sum_{i=1}^{n} (y_i - \hat{y}_{-i})^2$
- $\hat{y}_{-i}$: پیش بینی برای iامین داده، با استفاده از همه‌ی داده‌ها به‌جز iامین داده.
- عملکرد مشابهSSE دارد، مدلی بهتر است که PRESSکمتری دارد.
- $R^2_{PRESS} = 1 - \frac{PRESS}{SST}$

- **معیارهایی که به تفاوت بین توزیع واقعی و توزیعی که توسط زیرمدل منتخب، مشخص شده می‌پردازند**

- اگر f(y) توزیع مدل واقعی باشد و g(y) توزیع مشخص شده توسط مدل داوطلب باشد، میزان اختلاف بین این دو توزیع به صورت زیر است

$$KL(f,g) = \int log\frac{f(y)}{g(y)} * f(y)\, dy$$

به طوریکه

$$KL(f,g) \neq KL(g,f)$$

یک مقدار واقعی است از اختلاف بین مدل برازش واقعی f و مدل رقیب g

$$KL(f,g) \geq KL(f,f) = 0$$

هرچه مقدار آماره KL نزدیک به صفر باشد، مدل کاندید شده بهتر است.

- معیارهایی که به تفاوت بین توزیع واقعی و توزیعی که توسط زیرمدل منتخب مشخص شده می‌پردازند

- **AIC: Akaike Information Criterion**

- برای مدل‌های احتمالی به‌وسیله‌ی تقریب متوسط زیان استفاده می‌شود که توسط Akaike در سال 1974 مطرح شد.

$$AIC = nLog\left(\frac{SSE_p}{n}\right) + 2p$$

$$\text{AIC} = -2LogL(M) - 2p(M)$$

که L تابع ماکزیمم درستنمایی که با جایگذاری MLEحاصل می‌شود و p تعداد پارامتر مدل موردنظر است.

- minimize these for “best model”.

- معیارهایی که به تفاوت بین توزیع واقعی و توزیعی که توسط زیرمدل منتخب مشخص شده می‌پردازند

- FPE_α: Final Prediction Error

- برای مدل‌های رگرسیون خطی به‌وسیله‌ی براورد خطای پیش بینی مورد استفاده قرار می‌گیرد که توسط Akaike در سال 1969 مطرح شد.

$$\mathrm{FPE}_\alpha = n\hat{\sigma}_k^2 + \alpha k s_K^2 = \mathrm{RSS}_k + \alpha k s_K^2,$$

$$\text{where } s_k^2 = \mathrm{RSS}_k/(n-k),$$

minimize these for “best model”

معیارهایی که احتمال پسین را تخمین می‌زنند

BIC: Bayesian Information Criterion

به‌وسیله‌ی تقریب احتمال پسین، مدل موردنظر به‌دست می‌آید که توسط Schwarz در سال 1978 مطرح شد.

$$BIC(M) = nLog\left(\frac{SSE_p}{n}\right) + pLog(n)$$

$$BIC(M) = -2LogL(M) + p(M)Log(n)$$

که L تابع ماکزیمم درستنمایی که با جایگذاری MLEحاصل می‌شود و p تعداد پارامتر مدل موردنظر است.

- minimize these for “best model”.

- **معیارهای تعمیم یافته انتخاب مدل**

- AIC_α و FPE_α
- در گروه خانواده معیار اطلاعات تعمیم یافته **GIC** (Generalized Information Criterion)تعریف می‌شوند.

$$\mathrm{GIC} = -2\log\mathcal{L}_k + \alpha c_k,$$

- در یک مدل خطی GIC به این فرم است:

$$\mathrm{GIC} = n\log\hat{\sigma}_k^2 + \alpha k,$$

- معیار تعمیم یافته FPE_α جز خانواده GIC قرار دارد، همانطور که اشاره شد به فرم زیر است:

$$\mathrm{FPE}_\alpha = n\hat{\sigma}_k^2 + \alpha k s_K^2 = \mathrm{RSS}_k + \alpha k s_K^2,$$

✓زمانی‌که $\frac{\alpha k}{n} \rightarrow 0$

$$\frac{\log\{\mathrm{FPE}_\alpha/n\}}{\mathrm{GIC}/n} = \frac{\log\hat{\sigma}_k^2 + \log\{1+(\alpha k/n)(s_K^2/\hat{\sigma}_k^2)\}}{\log\hat{\sigma}_k^2 + \alpha k/n} \xrightarrow{a.s.} 1.$$

✓ برای انتخاب مدل‌های خطی به جای GIC از FPE_α استفاده می کنیم.

✓اگر در $\mathrm{FPE}_\alpha = \mathrm{RSS}_k + \alpha c_k s_K^2$، $c_k = (e^{\alpha k/n} - 1)\mathrm{RSS}_k/\alpha s_K^2$ باشد ،

FPE_α دقیقا برابر با AIC_α است.

❖استفاده از بیز، یک خانواده تعمیم یافته از معیار اطلاعات بیزی در اختیار ما می گذارد که به این فرم است:

$$\mathrm{BIC}_\pi = -2\log \mathcal{L}_k + k\log n - 2\log \pi_k,$$

❖که π_k احتمال پیشین مدل، L تابع ماکزیمم درستنمایی که با جایگذاری MLEحاصل می‌شود و k تعداد پارامتر مدل موردنظر است.

❖یک خانواده از معیارهای اطلاعات بیزی با احتمال پیشین برنولی BIC_q نامیده می شود و عملکرد آن از BIC بهتر خواهد بود.

1. Joseph B.kadane; Nicole A . Lazar; Methods and Criteria for model selection
2. Alvin C. Rencher and G. Bruce Schaalje ; LINEAR MODELS IN STATISTIC.
3. Gerda Claeskens, Nils Lid Hjort Model Selection and Model Averaging
4. Xu, Changjiang, "Model Selection with Information Criteria" (2010). Electronic Thesis and Dissertation Repository. 46

سپاس از
توجه شما