Model Comparison

PSYC 573

University of Southern California

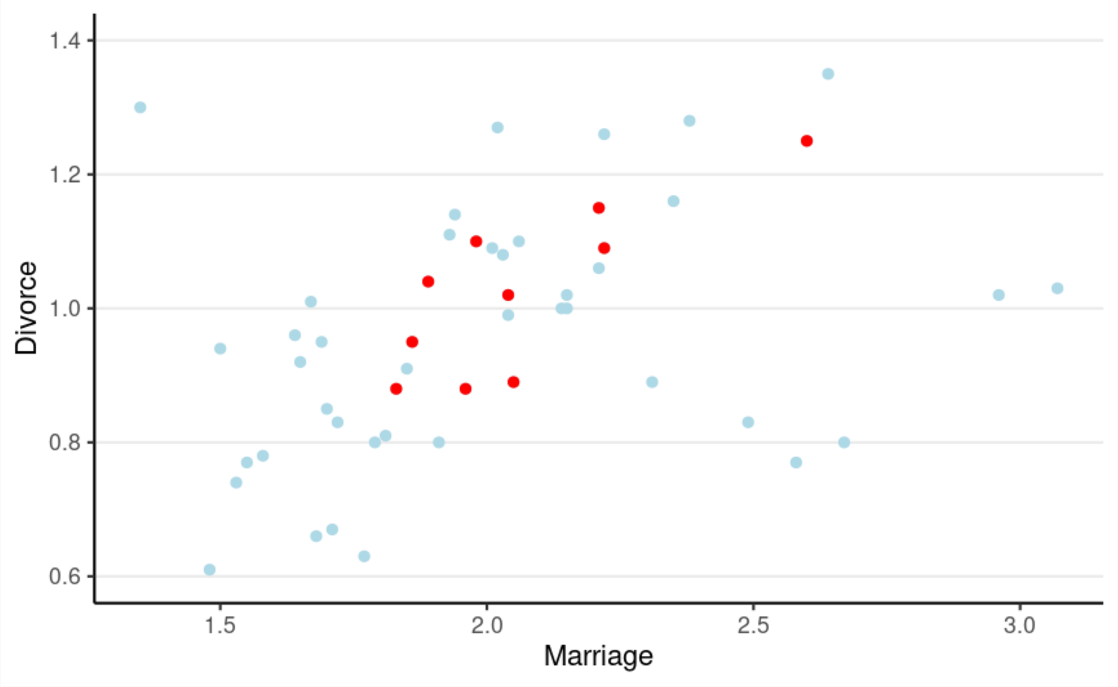
April 14, 2022

Guiding Questions

- What is *overfitting* and why is it problematic?
- How to measure *closeness* of a model to the true model?
  - What do *information criteria* do?

In-Sample and Out-Of-Sample Prediction

- Randomly sample 10 states

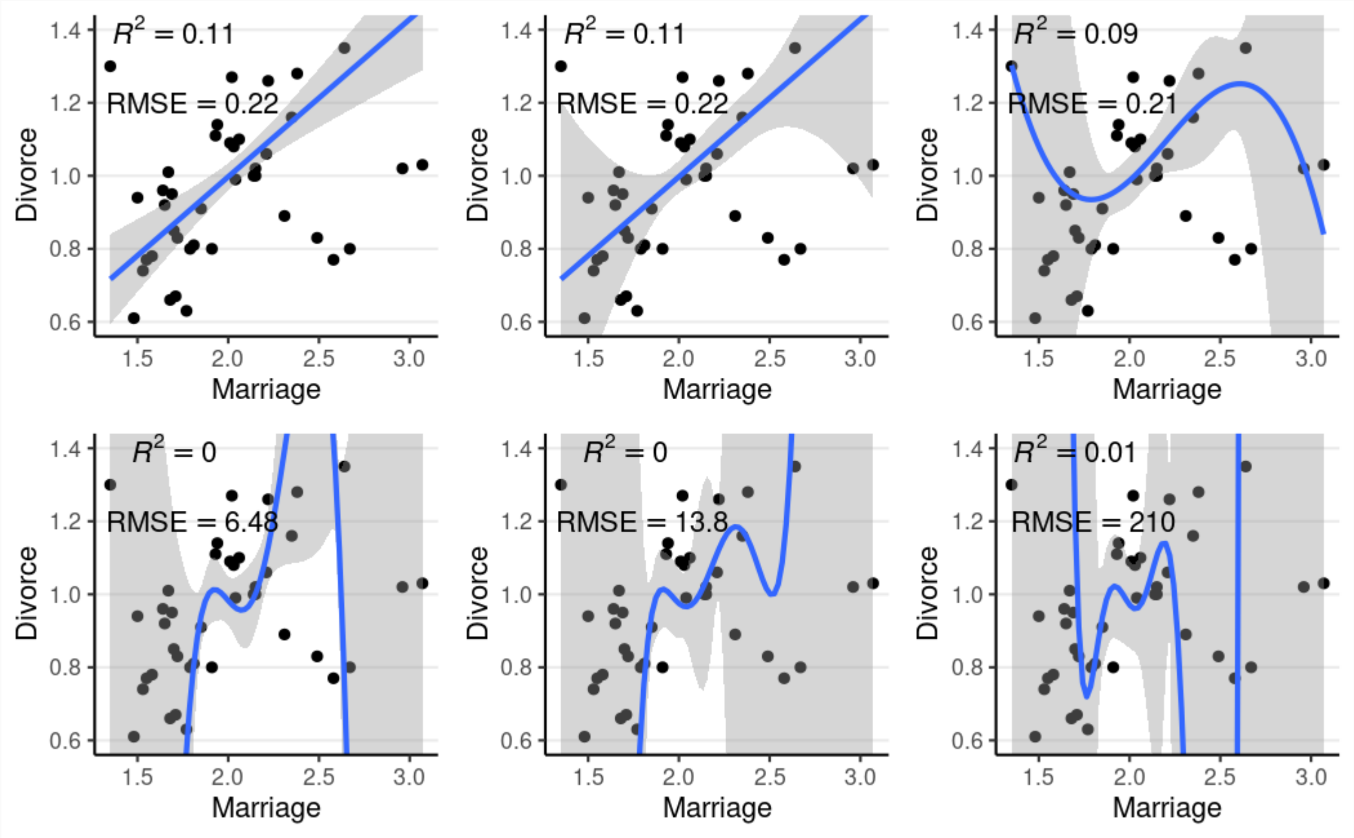


Underfitting and Overfitting

- Complex models require more data
  - Too few data for a complex model: **overfitting**
  - A model being too simple: **underfitting**

Prediction of Future Observations

- The more a model captures the noise in the original data, the less likely it predicts future observations well



What Is A Good Model?

- Closeness from the proposed model (M_1) to a "true" model (M_0)
  - *Kullback-Leibler Divergence* (D_{KL}) = $\textbf{Entropy of } M_0 - \textbf{elpd of } M_1$
  - elpd: expected log predictive density: $E_{M_0}[\log P_{M_1}(\tilde{\mathbf{y}})]$

- Choose a model with *smallest* D_{KL}
  - When $M_0 = M_1$, $D_{\text{KL}} = 0$
  - $\Rightarrow$ choose a model with largest elpd

Expected log *pointwise* predictive density

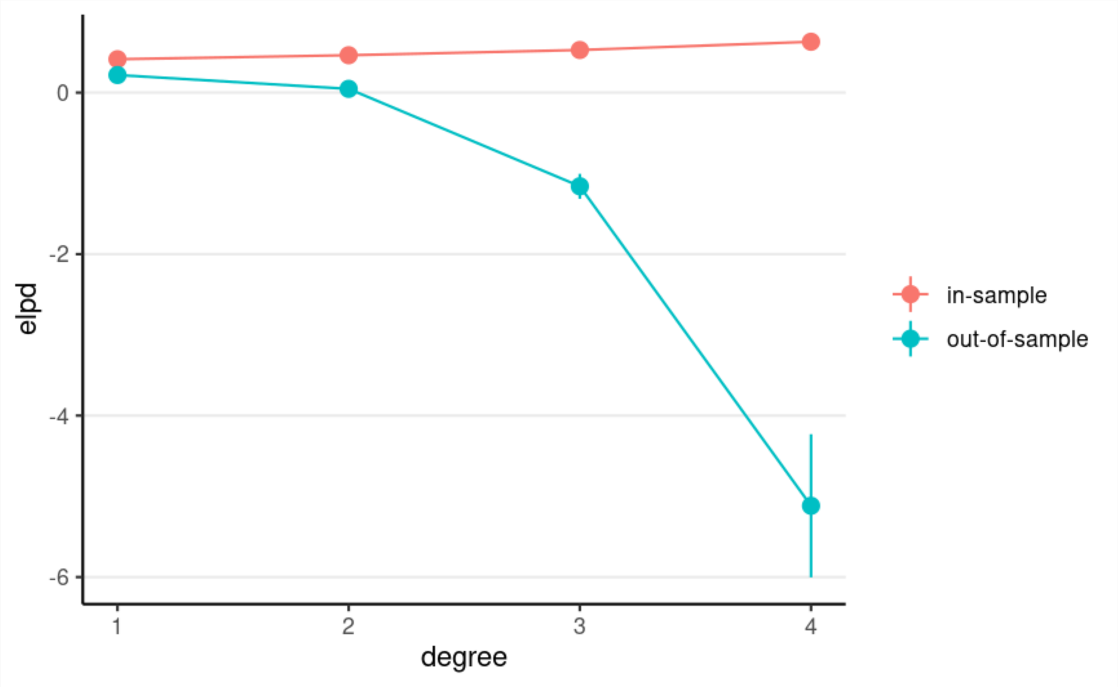
$$\sum_i \log P_{M_1}(y_i)$$

Note: elpd is a function of sample size

- Problem: elpd depends on M_0, which is unknown
  - Estimate elpd using the current sample $\rightarrow$ underestimate discrepancy
  - Need to estimate elpd using an *independent sample*

Overfitting

Training set: 25 states; Test set: 25 remaining states



- More complex model = more discrepancy between in-sample and out-of-sample elpd

Information Criteria (IC)

Approximate discrepancy between in-sample and out-of-sample elpd

IC = -2 $\times$ (in-sample elpd - $\boldsymbol{p}$)

$\boldsymbol{p}$ = penalty for model complexity

- function of number of parameters

Choose a model with **smaller** IC

Bayesian ICs: DIC, WAIC, etc

Cross-Validation

- Split the sample into K parts
- Fit a model with K - 1 parts, and obtain elpd for the "hold-out" part
- Very computationally intensive
- `loo` package: approximation using Pareto smoothed importance sampling

```r
loo(m1)
```

```
>#
># Computed from 8000 by 50 log-likelihood matrix
>#
>#           Estimate  SE
># elpd_loo     15.1 4.9
># p_loo         3.3 1.0
># looic       -30.2 9.9
># ------
># Monte Carlo SE of elpd_loo is 0.0.
>#
># All Pareto k estimates are good (k < 0.5).
># See help('pareto-k-diagnostic') for details.
```

Comparing Models

$$\texttt{\textbf{Divorce}}_i \sim N(\mu_i, \sigma)$$

- M1: `Marriage`
- M2: `Marriage`, `South`, `Marriage` $\times$ `South`
- M3: `South`, smoothing spline of `Marriage` by `South`
- M4: `Marriage`, `South`, `MedianAgeMarriage`, `Marriage` $\times$ `South`, `Marriage` $\times$ `MedianAgeMarriage`, `South` $\times$ `MedianAgeMarriage`, `Marriage` $\times$ `South` $\times$ `MedianAgeMarriage`

	M1	M2	M3	M4
b_Intercept	0.61	0.67	0.94	5.53
b_Marriage	0.18	0.13		−1.21
b_Southsouth		−0.62	0.10	0.32
b_Marriage × Southsouth		0.36		0.52
bs_sMarriage × SouthnonMsouth_1			−0.55	
bs_sMarriage × Southsouth_1			1.27	
sds_sMarriageSouthnonMsouth_1			0.91	
sds_sMarriageSouthsouth_1			0.48	
b_MedianAgeMarriage				−1.73
b_Marriage × MedianAgeMarriage				0.45
b_MedianAgeMarriage × Southsouth				−0.36
b_Marriage × MedianAgeMarriage × Southsouth				−0.08
ELPD	15.1	18.3	17.7	23.8
ELPD s.e.	4.9	5.5	5.8	6.1
LOOIC	−30.2	−36.6	−35.3	−47.5
LOOIC s.e.	9.9	11.0	11.7	12.1
WAIC	−30.3	−36.9	−37.1	−48.1
RMSE	0.17	0.15	0.14	0.13

Notes for Using ICs

- Same outcome variable and transformation
- Same sample size
- Cannot compare discrete and continuous models
  - E.g., Poisson vs. normal