Tuned penalized regression – a solution to the problem of separation in logistic regression?

Hana Šinkovec, Angelika Geroldinger, Rok Blagus, Georg Heinze

Our interests and methods

- the relationship between a binary response variable and other explanatory variables *X*

  Y=1 (occurrence of the response)

  Y=0 (non-occurrence)

- prediction of binary outcome → logistic regression

$$\Pr(\quad = 1) = \quad = [1 + \exp(- \quad)]^{-1}$$

- estimation of the parameters → maximum likelihood (ML)

$$\ell(\quad) = \log \quad (\quad) = \quad [\quad \log \quad + (1 - \quad)\log(1 - \quad)]$$

- exp(   ) = *odds ratio* should be interpretable

Separation

- under certain conditions:
  - small/sparse data set
  - rare outcomes/exposures
  - covariates with strong correlations/effects
- Example:

complete separation

	A	B
0	15	0
1	0	15

quasi-complete separation

	A	B
0	12	3
1	15	0

→ the two outcome groups are perfectly separated by the values of a covariate or a linear combination of covariates

ML parameter estimates:

$$= \log\left(\frac{_{11}\;\;_{22}}{_{12}\;\;_{21}}\right)$$

do not exist!

Penalized likelihood logistic regression

- intended to provide shrinkage of the parameter estimates → parameter estimates do not diverge

$$\ell\ (\ \ ) = \log\ (\ \ ) +\ \ (\ \ )$$

- Ridge (R): $(\ \ ) = -\ \ \ ^{2}$
- LASSO (L): $(\ \ ) = -\ \ |\ \ |$
- Firth (F): $(\ \ ) = \ \log|\ (\ \ )|$
- Log-*F*(*m*,*m*) (LF): $(\ \ ) = \log(|\ (\ \ )|^{-}\ \ )$

Tuning

- Cross-validation (CV) minimizes the prediction error of the model = CV deviance:

$$= -2\{ \quad \log \quad + (1 - \quad)\log(1 - \quad)\}$$

- Akaike information criterion (AIC):

$$= \quad [\quad \log \quad + (1 - \quad)\log(1 - \quad)] + 2 \quad =$$

$$-2\log \quad + 2 \quad = \quad + 2$$

$$= \quad + 2$$

Yes!

But ...

... the optimal penalty parameter is often zero leading to the standard maximum likelihood solution.

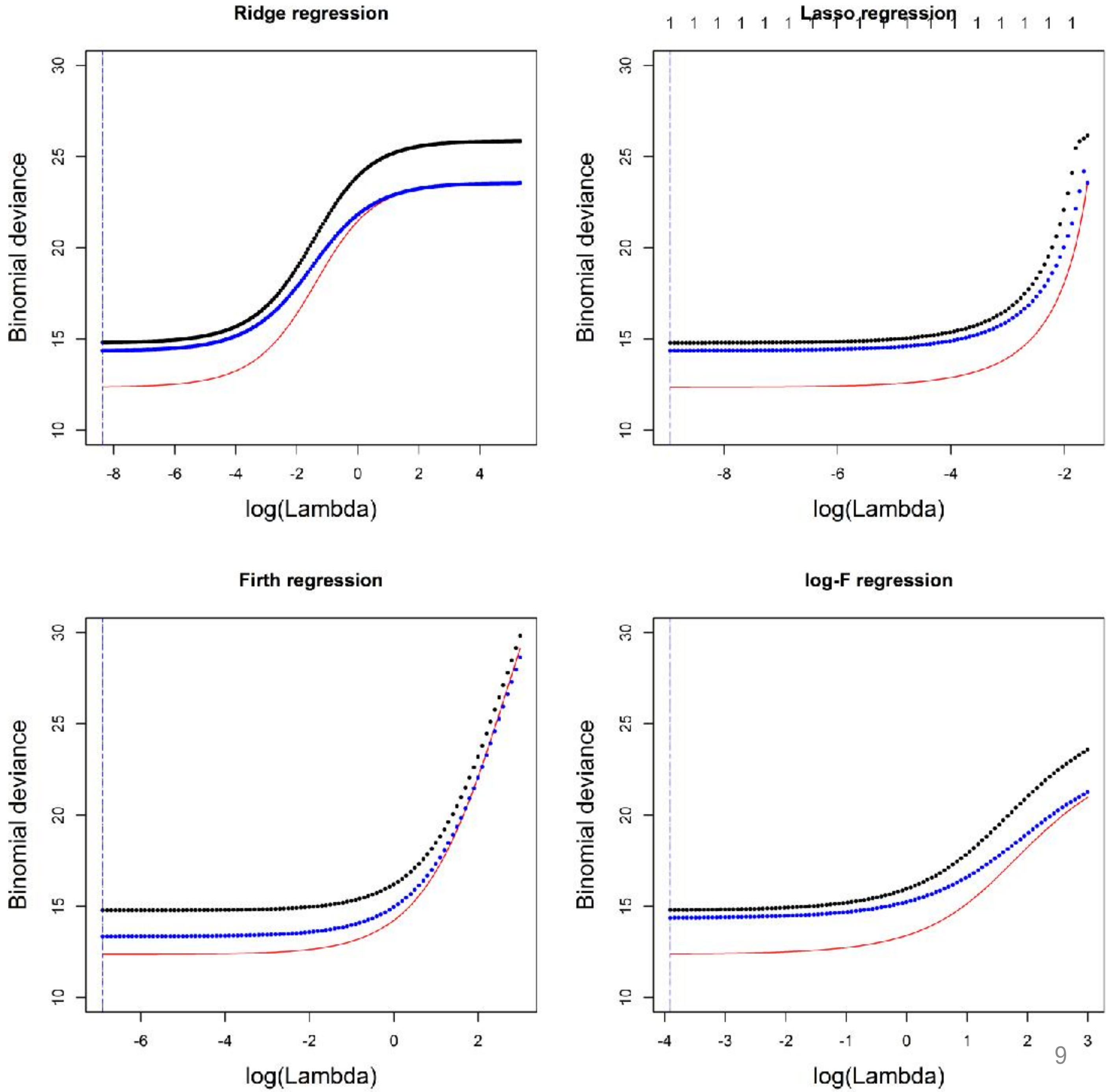
Has anyone of you ever run into this problem?

Constructed example

Settings:

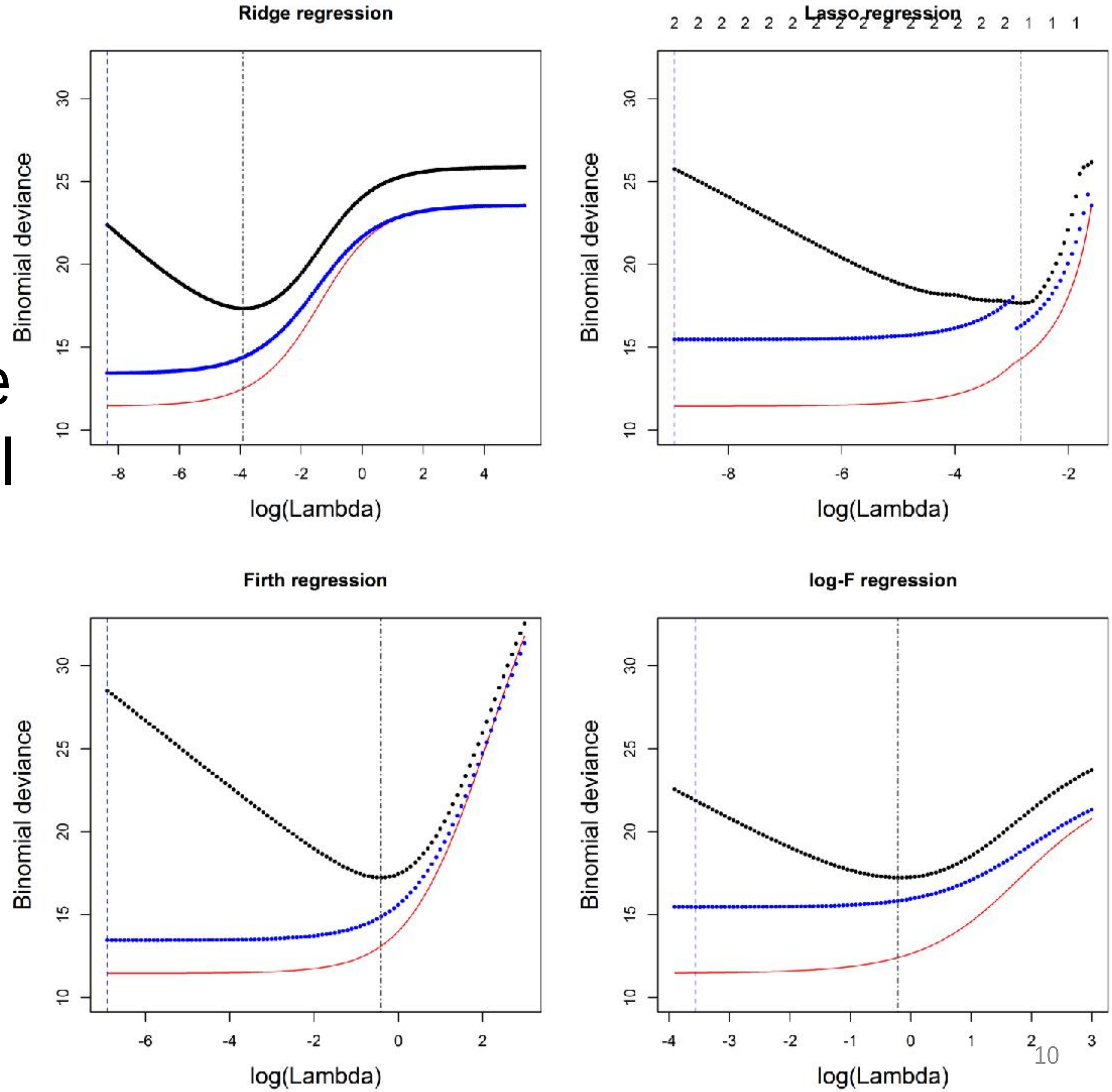
- $= 30$
- $_1 = 3, \quad _2 = 0$
- $(\quad) = 0.30, \quad (\quad) = 0.20$
- no corrlation between $_1$ and $_2$

	$_1 = 0$	$_1 = 1$
0	21	5
1	0	4

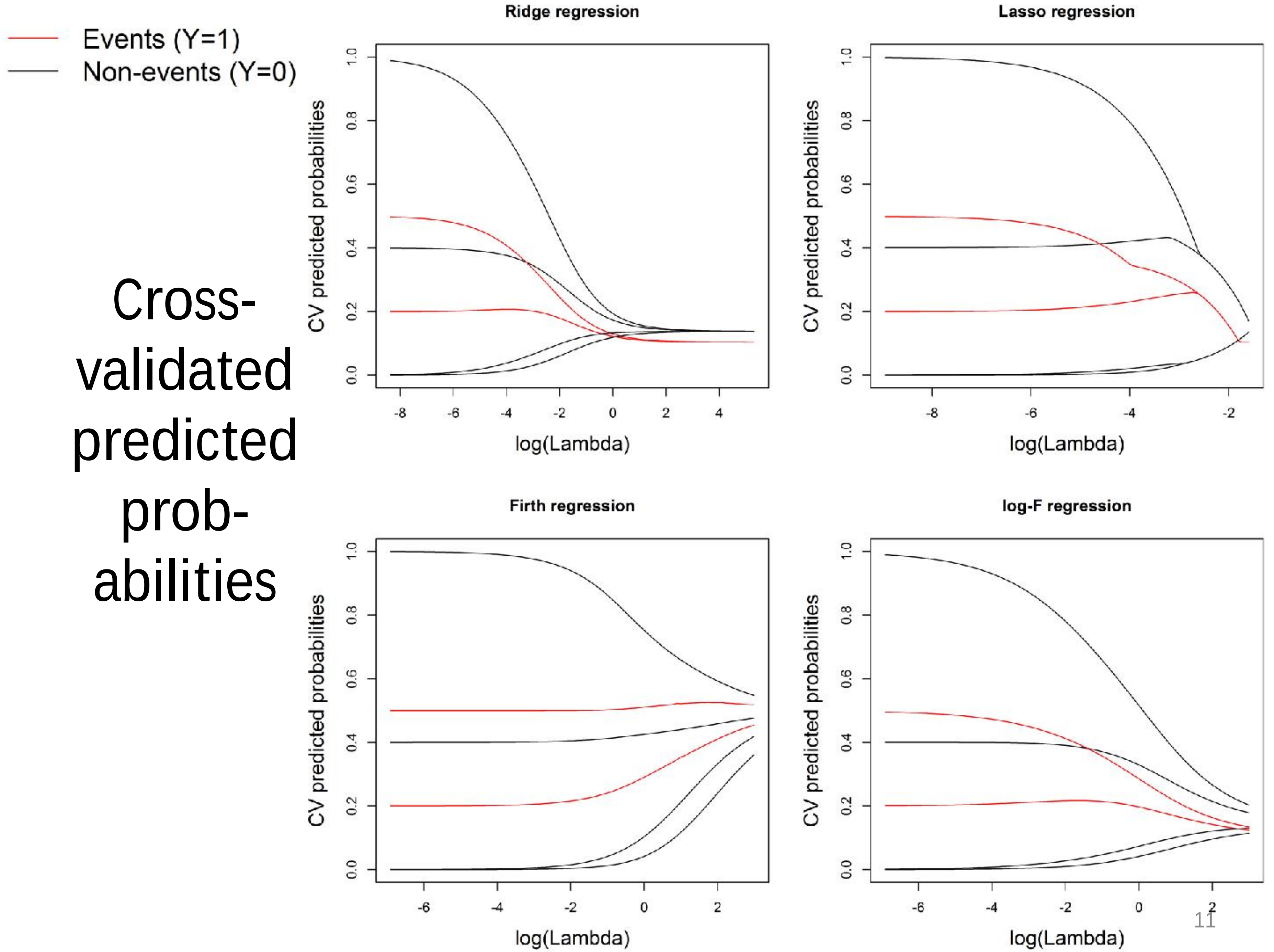


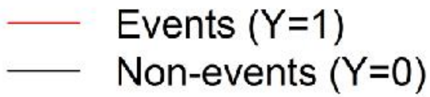
Let`s tune the model with 1

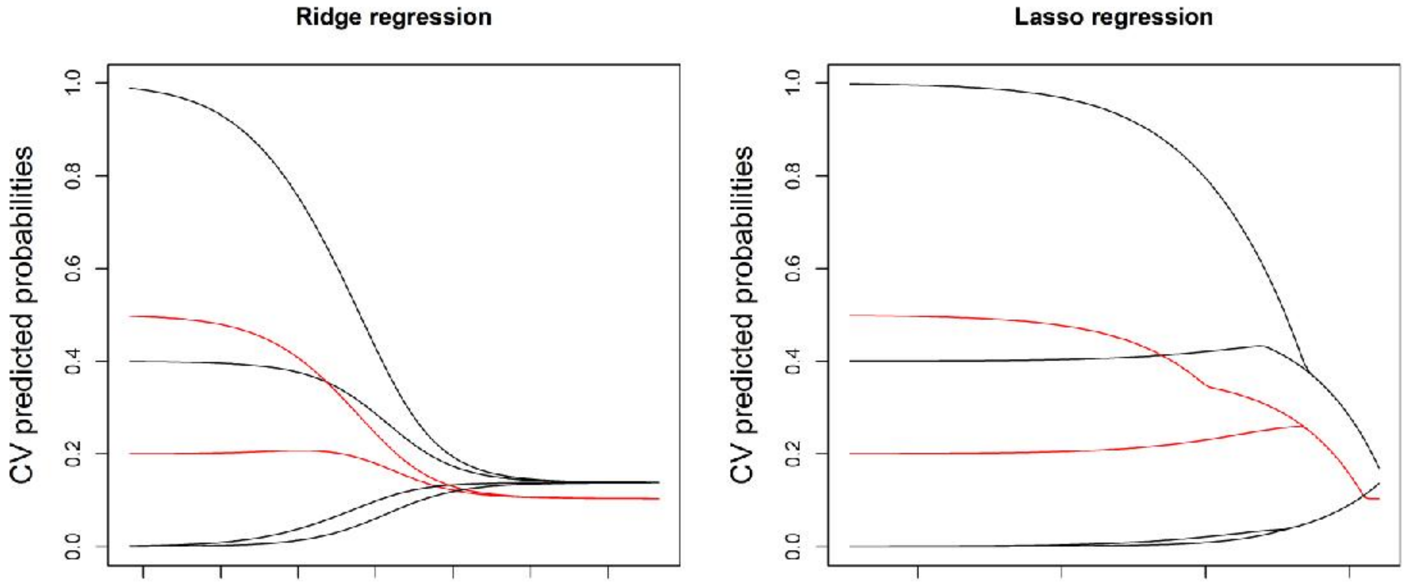
Let`s tune the model with 1 AND 2



Cross-validated predicted prob-abilities

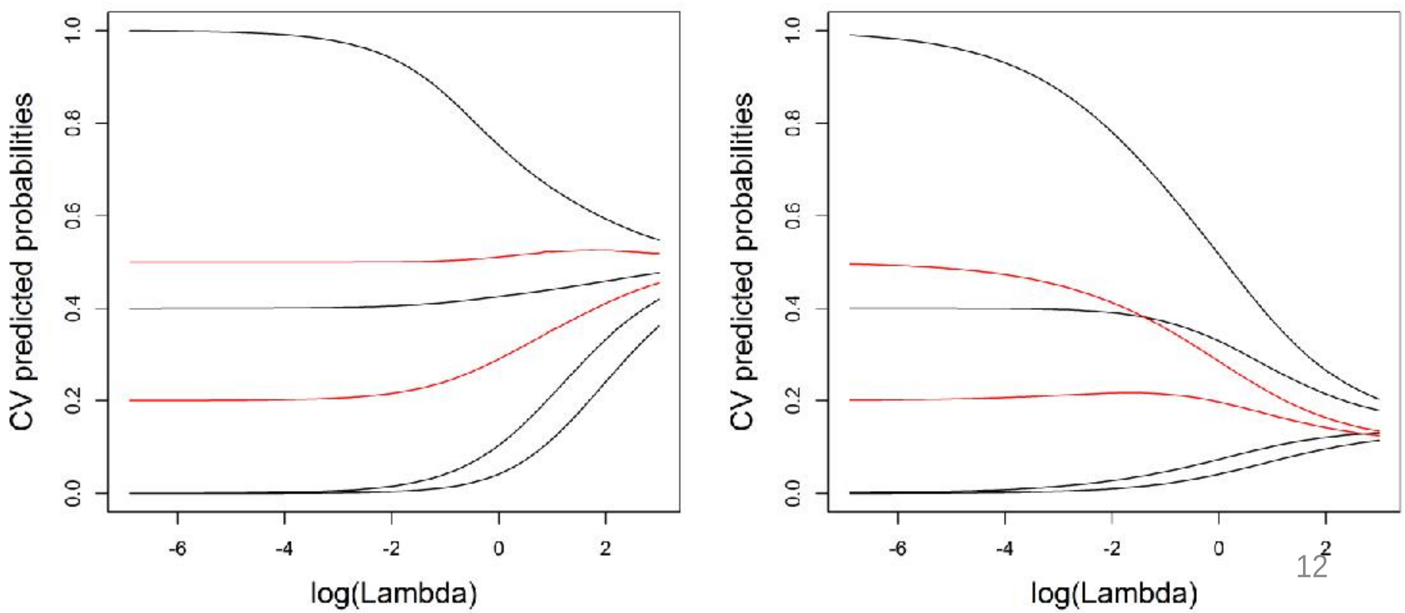






Cross-validated predicted prob-abilities

$$= -2\{ \quad \log \quad + (1 - \quad)\log(1 - \quad)\}$$



s obtained

	ML (U)	ML (M)	R (CV)	L (CV)	F (CV)	LF (CV)	R (AIC)	L (AIC)	F (AIC)	LF (AIC)
($_0$)		$-\infty$	−4.23	−3.23	−3.66	−4.33	$-\infty$	$-\infty$	$-\infty$	−7.90
$_1$	∞	∞	3.47	2.63	3.11	3.62	∞	∞	∞	7.21
$_2$	0.99	1.39	0.97	0.00	1.02	0.90	1.39	1.39	1.39	1.36

- ML (U) – univariable maximum likelihood analysis
- ML (M) – multivariable maximum likelihood analysis
- R – ridge regression
- L – LASSO regression
- FL – Firth regression
- LF – log-*F* regression

To do list

- Adding noise predictors removes the problem → does this improve estimation?
- Tuning can be unstable → is it better to pre-specify the value of penalty parameter?

My questions to the others

- Have you ever encountered separation?
- How did you fix non-existence of finite maximum likelihood parameter estimates?
- What kind of odds ratios did you offer to the clinician?
- Have you ever used tuned penalized regression?
- Have you ever had troubles applying resampling methods when using logistic regression?
- Have you ever encountered separation in a mixed logistic regression model? What did you do?