# Learning penalties for change-point detection using max-margin interval regression

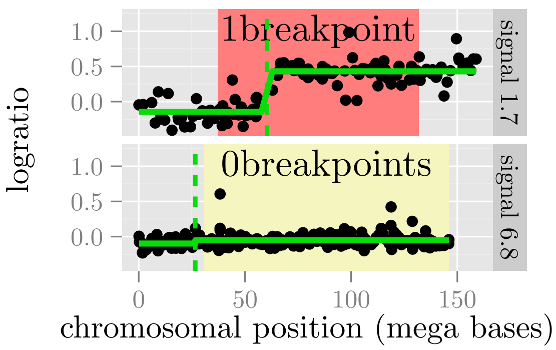
Toby Dylan Hocking and Guillem Rigaill

Acknowledgements
Gudrun Schleiermacher, Isabelle Janoueix-Lerosey, Francis Bach, Jean-Philippe Vert

15 October 2012

## Segmenting annotated signals

Given a signal $y \in \mathbb{R}^d$ (black points), find a segmentation $\hat{y} \in \mathbb{R}^d$ (green lines).




Annotation error $e_i : \{1, \dots, k_{\max}\} \to \mathbb{R}^+$ for model with $k$ segments

$$e_i(k) = \sum_{(r,a) \in (R_i, A_i)} 1_{|\hat{P}_i^k \cap r| \notin a}$$

for signal $i$ we have

- regions i.e. $R_i = \{[10, 20], [30, 40]\}$.
- annotations i.e. $A_i = \{\{0\}, \{1\}\}$.
- change-points $\hat{P}_i^k$ for model $k$, i.e. $\{25, 35\}$.

## Penalized model error

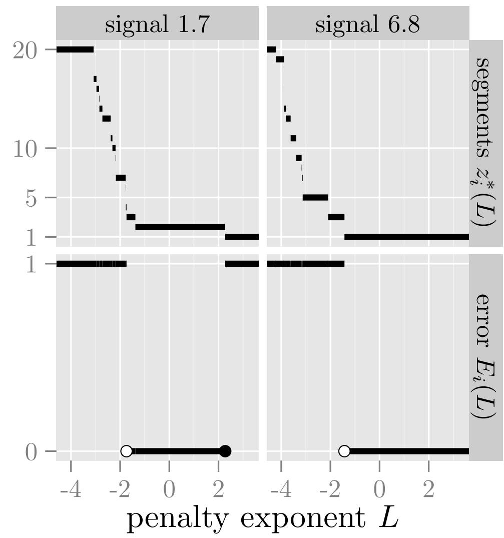
Number of segments $z_i^* : \mathbb{R} \to \{1, \dots, k_{\max}\}$

$$z_i^*(L) = \underset{k \in \{1,\dots,k_{\max}\}}{\arg\min} \exp(L)k + ||y_i - \hat{y}_i^k||_2^2$$

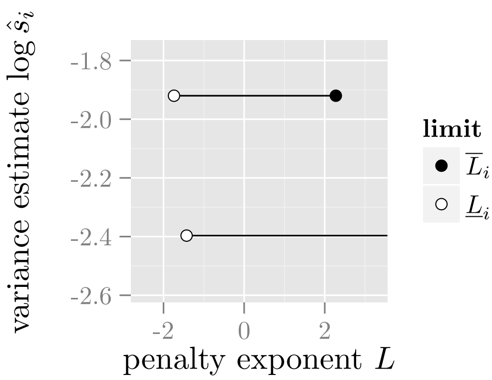
Penalized model annotation error $E_i : \mathbb{R} \to \mathbb{R}^+$

$$E_i(L) = e_i \left[ z_i^*(L) \right]$$

The $z_i^*$ and $E_i$ are piecewise constant functions that can be calculated exactly, and are plotted below for 2 signals $i$.



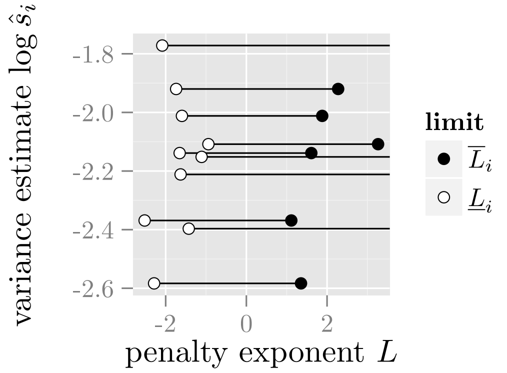

The target interval of the error curves above is plotted as a function of a variance estimate feature below.



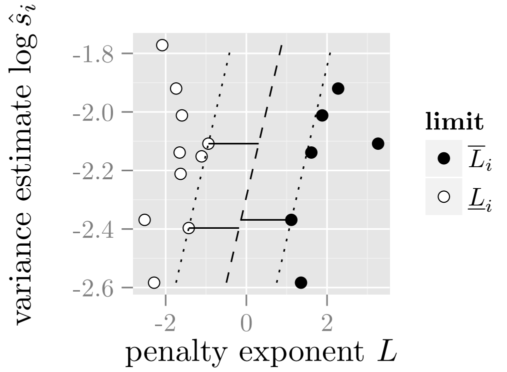

We will build a model that predicts model complexity $L$ from features such as the variance $\hat{s}$.

## The learning problem

Find a line that intersects every target interval.

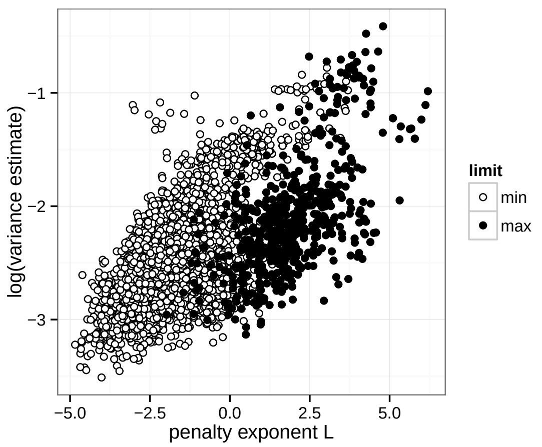



Or, a line that separates the limits.




There are infinitely many separators in the small data set above, so we take the line with maximum margin (horizontal black lines).

In the larger data below, there is no separator.




Source: `http://cran.r-project.org`.

```
library(neuroblastoma)
data(neuroblastoma)
```

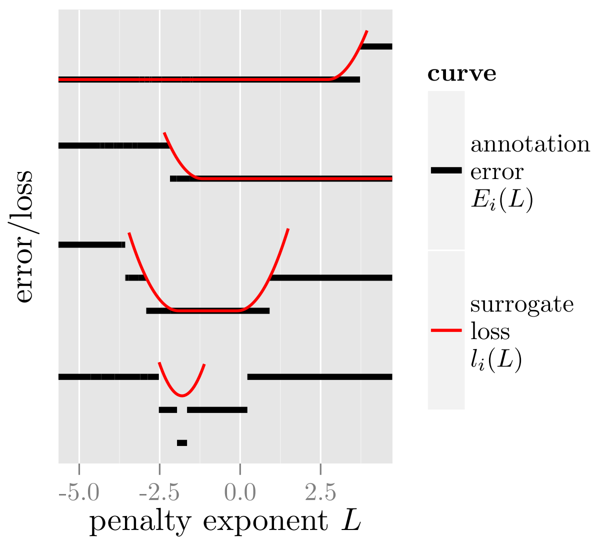
## A convex relaxation

The annotation error is non-convex

$$\arg\min_f \sum_{i=1}^n E_i[f(x_i)],$$

but we use a convex relaxation

$$\arg\min_f \sum_{i=1}^n l_i[f(x_i)].$$




We define the surrogate loss as

$$l_i(L) = \phi(L - \underline{L}_i) + \phi(\overline{L}_i - L)$$

where the squared hinge loss $\phi : \mathbb{R} \to \mathbb{R}^+$ is

$$\phi(L) = \begin{cases} (L-1)^2 & \text{if } L \leq 1 \\ 0 & \text{if } L \geq 1. \end{cases}$$

The resulting convex optimization problem

$$\arg\min_{\beta\in\mathbb{R}, w\in\mathbb{R}^m} \gamma||w||_1 + \frac{1}{n}\sum_{i=1}^n l_i(w'x_i + \beta)$$

for some fixed regularization parameter $\gamma \in \mathbb{R}^+$ can be solved using accelerated proximal gradient methods such as FISTA [BT09].

## References


[BT09] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2(1):183–202, 2009.