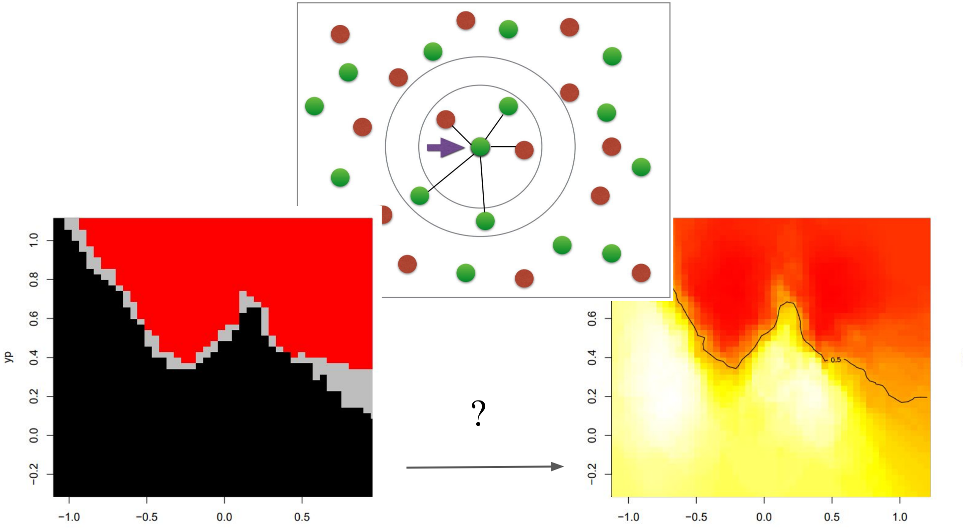
A Bayesian reassessment of nearest-neighbour classification

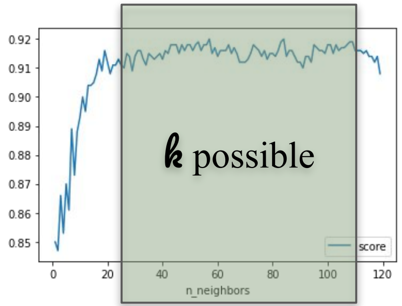
Is it possible to base the k-nearest neighbor model on a probabilistic model ?

By Lionel Cucala & al. (2009)



Etienne Peyrot
Charbel-Raphaël Segerie
MVA Bayesian Machine Learning

Probabilistic Framework



Why probabilistic ? Consistent model + Multiple ***k*** possible

How probabilistic ?

Potts Model : count the number of ***k***-neighbors ***ℓ*** with the same color as ***i*** :

$$\sum_{\ell \sim_k i} \delta_{y_i}(y_\ell)$$

Boltzmann probability :

$$f(\mathbf{y}|\mathbf{X}, \beta, k) = \exp\left(\beta \sum_{i=1}^{n} \sum_{\ell \sim_k i} \delta_{y_i}(y_\ell) \Big/ k\right) \Big/ Z(\beta, k)$$

Z is hard to compute !

The goal of the game: find the posterior distribution of beta and k through HM

How to win ? Succeed in bypassing the normalization constant Z(β, k)

Approach 1 : Pseudo Likelihood

indep.

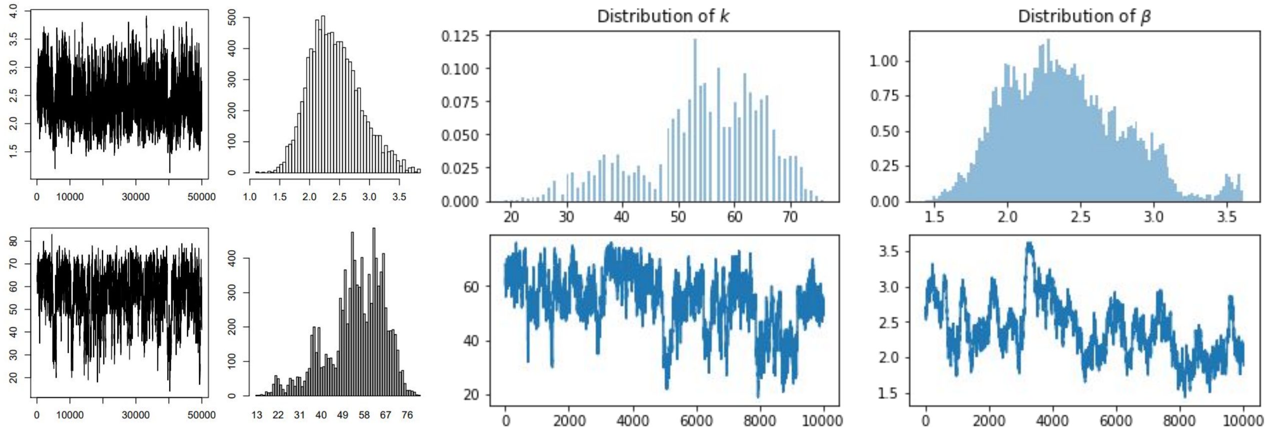
$P(y_i \mid y_{-i}, X, \beta, k)$

Estimation of the likelihood :

$$\hat{f}(\mathbf{y}|\mathbf{X}, \beta, k) = \prod_{i=1}^{n} \frac{\exp\left\{\beta/k\left(\sum_{\ell \sim_k i} \delta_{y_i}(y_\ell) + \sum_{i \sim_k \ell} \delta_{y_\ell}(y_i)\right)\right\}}{\sum_{g=1}^{2} \exp\left\{\beta/k\left(\sum_{\ell \sim_k i} \delta_{g}(y_\ell) + \sum_{i \sim_k \ell} \delta_{y_\ell}(g)\right)\right\}}$$

Normally $Z(\beta, k)$ = sum over 2^n states.

β



k

Approach 2 : True Likelihood / Path Sampling

$$S(\mathbf{y}) = \sum_i \sum_{\ell \sim_k i} \delta_{y_i}(y_\ell)/k$$

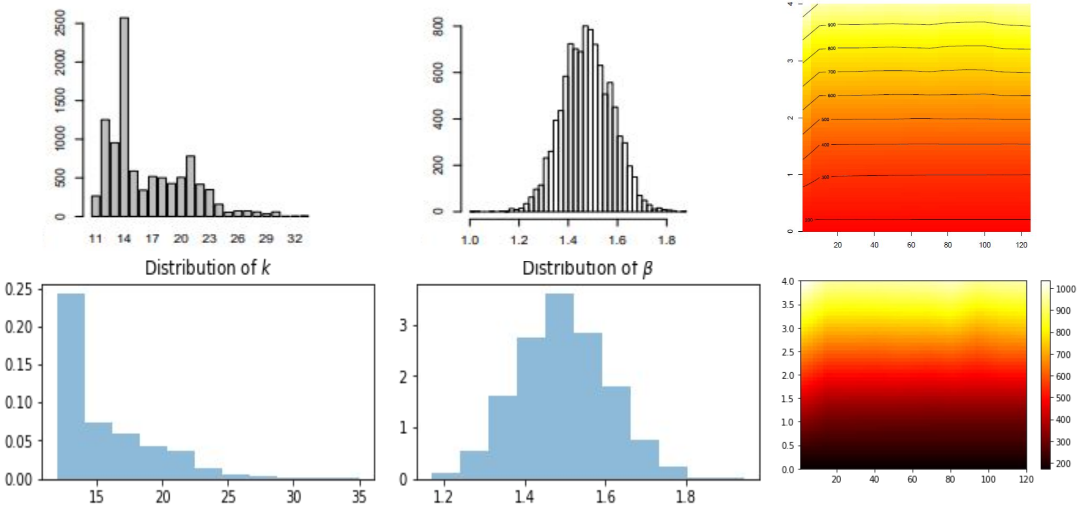
log Z(0, k)

$$Z(\beta, k) = \sum_{\mathbf{y}} \exp\left[\beta S(\mathbf{y})\right]$$

Using this equality :

$$\log\left(Z(\beta, k)\right) = n \log 2 + \int_0^\beta \mathbb{E}_{u,k}[S(\mathbf{y})]\, \mathrm{d}u$$

► Estimate Z(β, k) with a Monte Carlo sum



Approach 3 : Perfect Sampling

Add an arbitrary variable z :

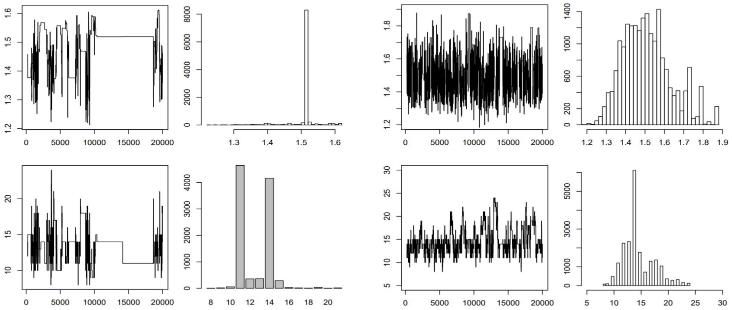
$$\pi(\beta, k, \mathbf{z}|\mathbf{y}) \propto \pi(\beta, k, \mathbf{z}, \mathbf{y}) = g(\mathbf{z}|\beta, k, \mathbf{y}) \times f(\mathbf{y}|\beta, k) \times \pi(\beta, k)$$

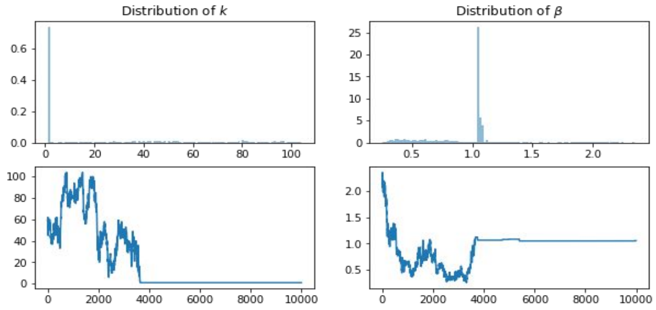
$$q_2(\beta', k', \mathbf{z}'|\beta, k, \mathbf{z}) = q_1(\beta', k'|\beta, k, \mathbf{y}) f(\mathbf{z}'|\beta', k'),$$

Metropolis-Hastings acceptance probability:

$$\left(\frac{\cancel{Z(\beta, k)}}{\cancel{Z(\beta', k)}}\right)\left(\frac{\overbrace{\exp\left(\beta' S(\mathbf{y})/k'\right)}^{P(y \mid \beta', k')}\,\overbrace{\pi(\beta', k')}^{P(\beta, k)}}{\exp\left(\beta S(\mathbf{y})/k\right)\pi(\beta, k)}\right)\left(\frac{\overbrace{g(\mathbf{z}'|\beta', k', \mathbf{y})}^{P(z \mid \beta, k, y)}}{g(\mathbf{z}|\beta, k, \mathbf{y})}\right)$$

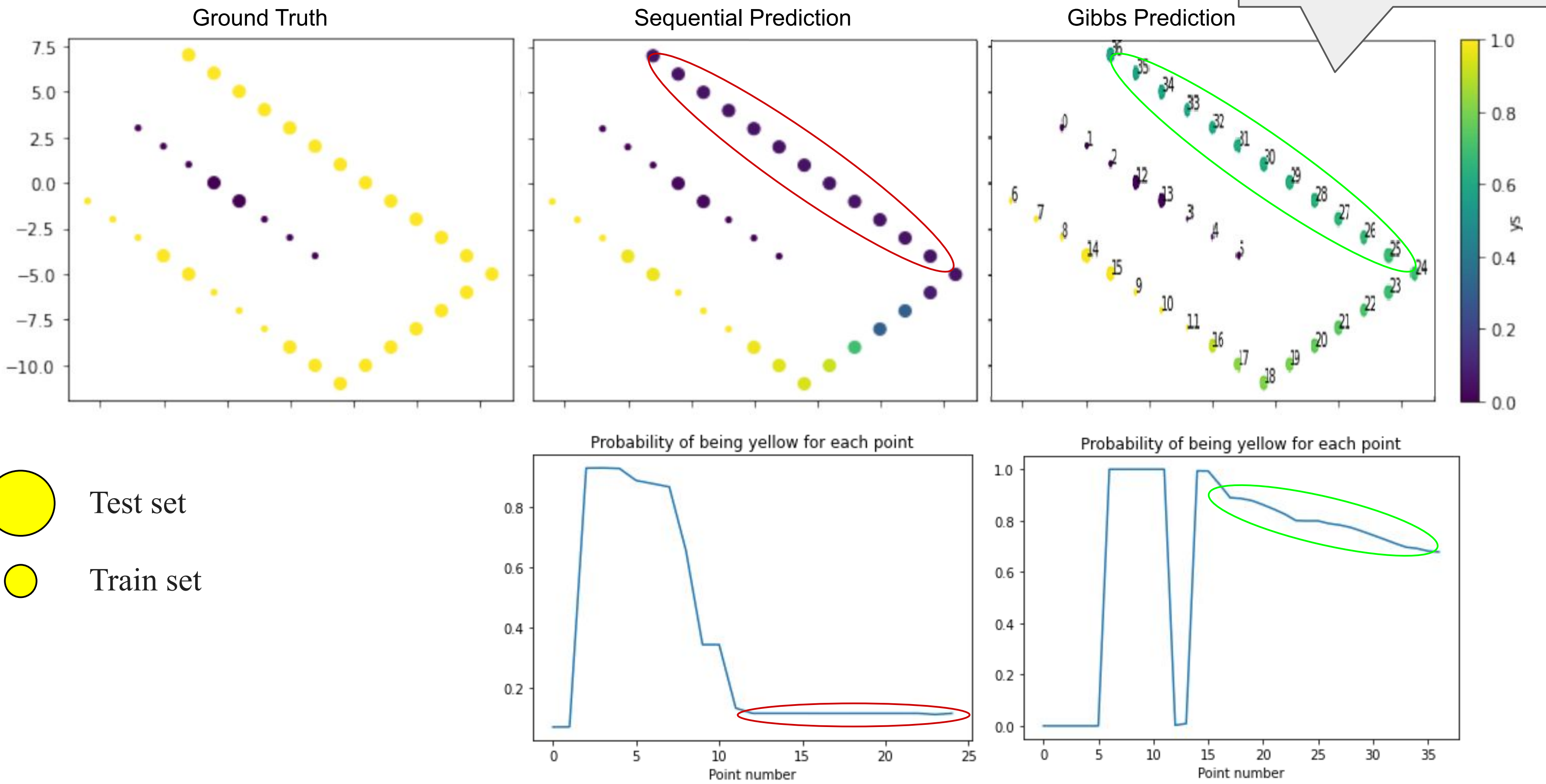
$$\times\left(\frac{q_1(\beta, k|\beta', k, \mathbf{y})\exp\left(\beta S(\mathbf{z})/k\right)}{\underbrace{q_1(\beta', k'|\beta, k, \mathbf{y})}_{P(\beta', k' \mid \beta, k)}\,\underbrace{\exp\left(\beta' S(\mathbf{z})/k'\right)}_{P(z' \mid \beta', k')}}\right)\left(\frac{\cancel{Z(\beta', k')}}{\cancel{Z(\beta, k)}}\right)$$





Our contribution : Gibbs Prediction

Our Gibbs prediction manages to extend the yellow structure all the way up!

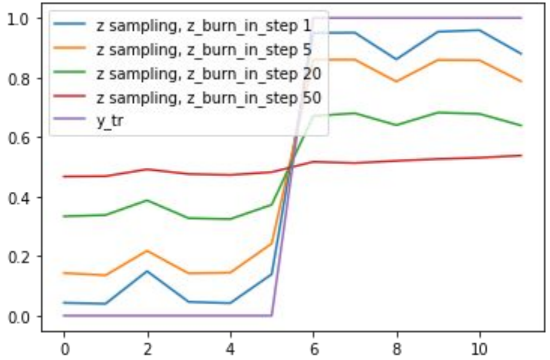


Test set

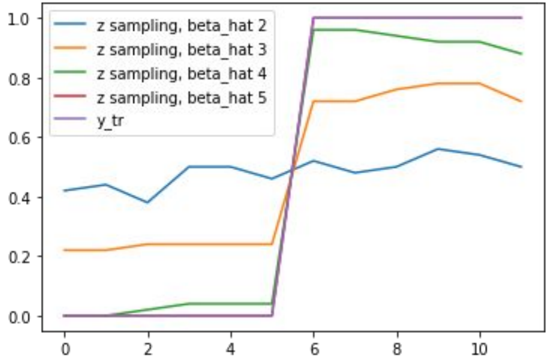
Train set

Bonus : Meta-parameters exploration

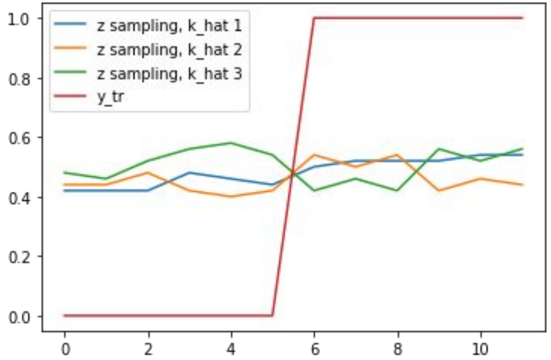
Burn in steps



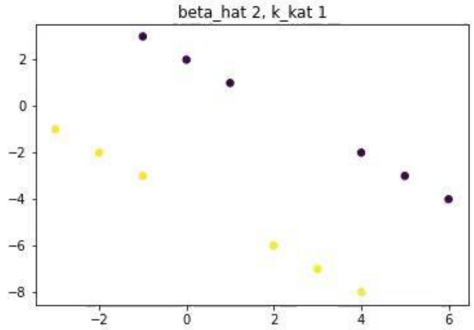
β



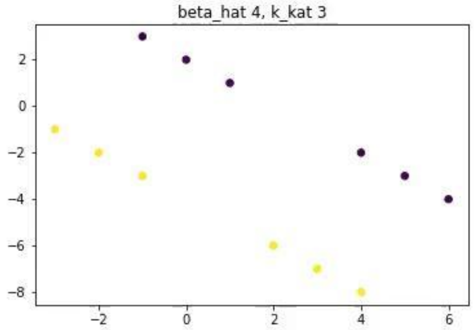
k



typical exploration



Big β : no exploration



Small β : uniform distribution