Multisolution Approach to Classification Tasks in Biomedicine

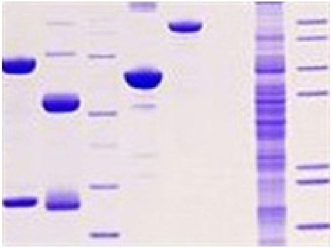
Kateřina Henclová

FNSPE CTU

June 27, 2021

Inspiration by Applications: Biomedicine

- Analyzed blood samples
- Find biomarkers of a disease
- Predict disease



The Classification/Regression/Feature Selection Problem

- Few samples
- $10^2 - 10^4$ features
- Uncertain data
- Interpretability required
- Multiple solutions possible
- Allow a domain expert to choose

	disease?	feature 1	feature 2	feature 3	feature 4	feature 5	feature 6	feature 7	feature 8	feature 9	feature 10	feature 11	feature 12
patient 1	yes	1036,9	364,9	777,7	2736,7	583,2	2052,5	4,37	1539,4	17,5	1154,6	54,0	865,9
patient 2	yes	1395,6	412,5	1046,7	3093,7	785,0	2320,3	5,89	1740,2	23,6	1305,2	64,8	978,9
patient 3	yes	1885,6	6533,4	1414,2	4900,0	1060,6	183,8	7,95	137,8	31,8	103,4	685,2	77,5
patient 4	yes	1948,3	5081,5	1461,2	3811,2	1095,9	142,9	8,22	107,2	32,9	80,4	541,0	60,3
patient 5	no	678,7	10945,3	509,0	820,9	381,8	615,7	2,86	461,8	11,5	346,3	1106,0	259,7
patient 6	yes	4320,9	14755,7	3240,7	3688,9	2430,5	2766,7	9,23	2075,0	36,9	1556,3	1512,5	1167,2
patient 7	yes	1805,7	3493,6	1354,3	2620,2	1015,7	1965,2	7,62	1473,9	30,5	1105,4	379,8	829,1
patient 8	no	178,5	5588,4	133,9	419,1	100,4	314,3	0,75	235,8	3,0	176,8	561,9	132,6
patient 9	no	400,2	3186,6	300,2	298,7	225,1	224,1	1,69	168,0	6,8	126,0	325,4	94,5
patient 10	no	590,5	1951,2	442,8	1363,4	332,1	1022,6	2,49	766,9	10,0	575,2	205,1	431,4

The Problem

$$y = X\beta \ (+ \text{ noise})$$

Analytical view: find βs that are both sparse and accurate:

$$||X\beta - y||_{l_2} \leq \varepsilon \quad \text{and} \quad ||\beta||_{l_0} \leq \delta.$$

Probabilistic view: what βs are most likely given data X, y?

Assume: noise $\sim \mathcal{N}(0, \sigma_e^2 I) \implies$ Maximum likelihood estimate:

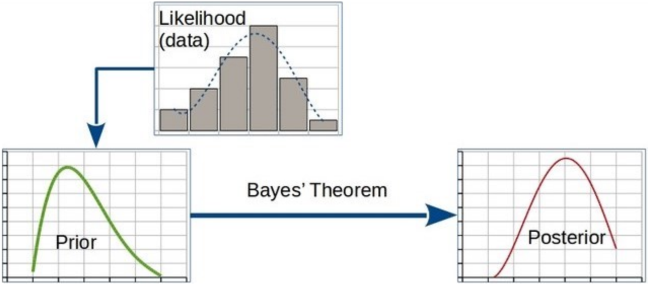
$$\beta_{MLE} = \text{argmax}_\beta \log \text{p}(y|\beta) = \text{argmin}_\beta ||X\beta - y||_{l_2}^2.$$

Each valid solution β corresponds to a peak in likelihood.

Approach to the Problem

1. Bayes: admit uncertainty
2. Approximation
3. New optimization problem
4. Stochastic gradient descent

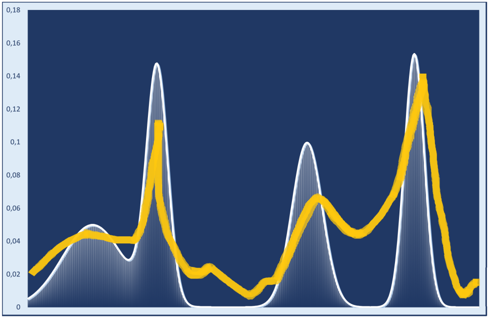
Bayesian Inference



$$\text{posterior distribution} = \frac{\text{evidence from data} \times \text{prior distribution}}{\text{normalization constant}}$$

Posterior Approximation by Gaussian Mixtures

- True posterior: too difficult to compute
- Approximation: multimodal, easier to compute
  (use mixture of Gaussians ⇒ find its parameters)
- Goodness of fit measure: Kullback-Leibler divergence
  (trick: Evidence Lower Bound ⇒ solve a different problem)



The Problem to Be Computed

Definition (The Parametrized Maximization Problem)

$$\operatorname{argmax}_{\forall \mu^{(k)},\ \forall \Sigma^{(k)},\ \forall \alpha_k} \iint q(\mathcal{Z}) \log \left(\frac{p(\mathcal{X}|\mathcal{Z})\ p(\mathcal{Z})}{q(\mathcal{Z})} \right) \ d\mu\ d\sigma,$$

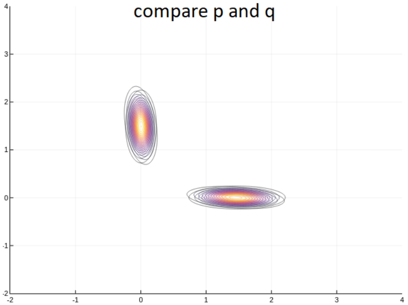
where:

- $\mathcal{X} = \{X, y\}$ ... all known stuff,
- $\mathcal{Z} = \left\{\mu^{(k)},\ \Sigma^{(k)},\ \alpha_k,\ \dots\right\}$ ... all unknown stuff,
- $p(\mathcal{Z})$ ... appropriate sparse prior,
- $p(\mathcal{X}|\mathcal{Z})$ ... data likelihood ($y \sim \mathcal{N}\left(X\mu^{(k)}, \sigma_e^2 I\right)$),
- $q(\mathcal{Z}) = \sum_{k=1}^{m} \alpha_k \mathcal{N}(\mu^{(k)}, \Sigma^{(k)})$ ... posterior's approximation.

Solving the Optimization Problem

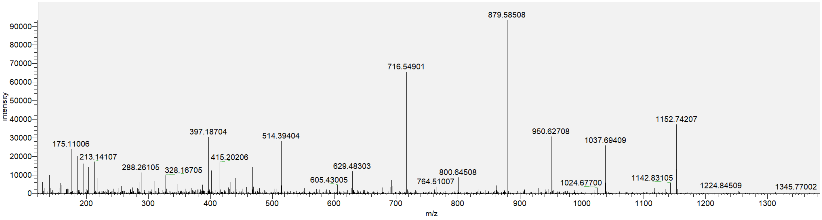
Use Stochastic Gradient Descent:

- Implicit reparametrization of mixture gradients
- Modified ADAM optimizer
- Automatic differentiation



Overview

- **Problem:** feature selection within classification/regression
- **Idea:** provide all solutions of required quality & sparsity
- **Approach:**
  1. Bayes theorem (sparse prior)
  2. Posterior approximation (multimodal Gaussian mixture)
  3. New optimization problem (use ELBO)
  4. Stochastic Gradient Descent (hardcore & tailored)



Thank you for your attention.