Extensions & summary

Sparsity of documents

$$p(W, Z, \Theta) = \prod_{d=1}^{D} p(\theta_d) \prod_{n=1}^{N_d} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn})$$

$$p(\theta_d) \sim \mathrm{Dir}(\alpha)$$

Aha. So we can allocate more topics for each document depending on \alpha. Why?

$\alpha \uparrow \quad \Rightarrow$ **More** topics for each document

$\alpha \downarrow \quad \Rightarrow$ **Less** topics for each document

α can be selected as $p(W|\alpha) \to \max\limits_{\alpha}$

W is the data. This is the maximum likelihood principle: ^^

Sparsity of topics

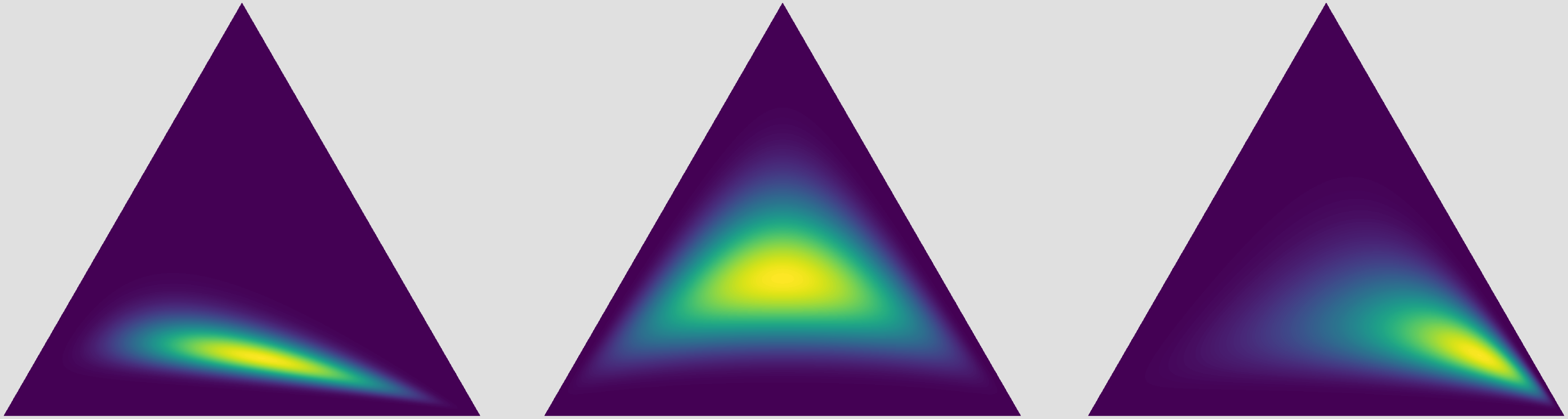
Sparse prior on Φ

$$p(W, Z, \Theta, \Phi) = \prod_{t=1}^{T} p(\Phi_{t\bullet}) p(W, Z, \Theta | \Phi)$$

$$p(\Phi_{t\bullet}) \sim \mathrm{Dir}(\beta)$$

Topics correlation

Logistic normal distribution



$$p(\theta_d) \sim \mathcal{P}(\mathcal{N}(\mu, \Sigma))$$

- Stars
- Astronomers
- Universe
- Galaxy

- Laser
- Optical
- Light
- Particles

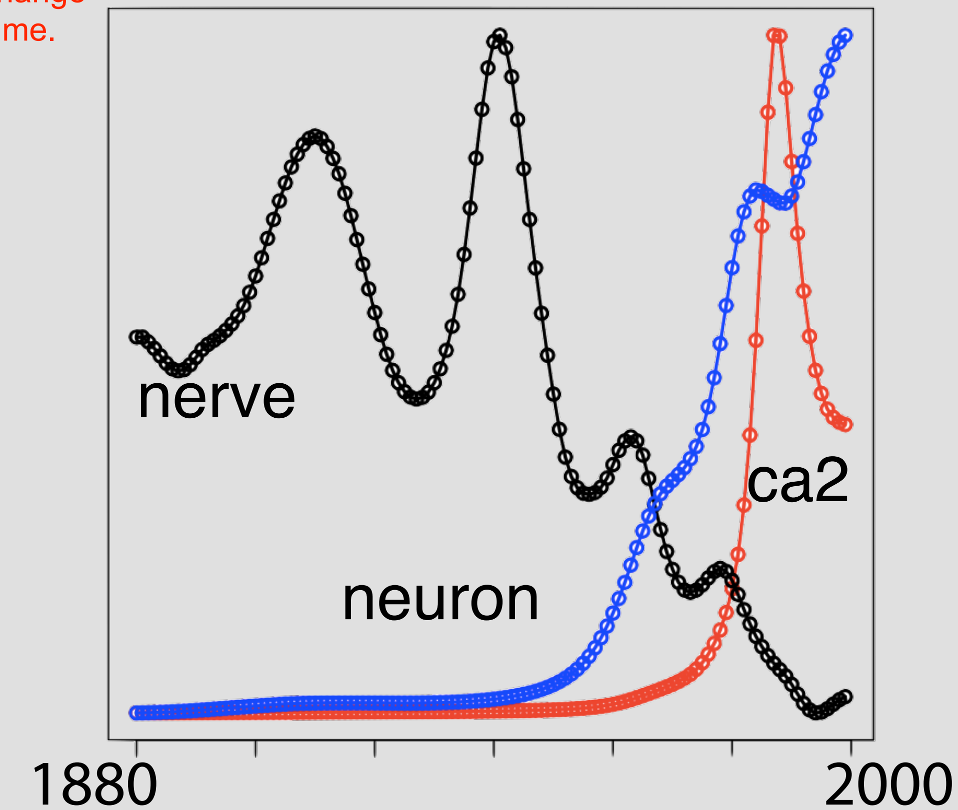
- Physics
- Particles
- Experiment
- Physicist

Dynamic Topic Model

$$p(B_{t\bullet}^{\tau+1}|B_{t\bullet}^{\tau}) \sim \mathcal{N}(B_{t\bullet}^{\tau}, \sigma^2 I)$$

$$\Phi_{t\bullet}^{\tau+1} = \mathrm{Softmax}[B_{t\bullet}^{\tau}]$$

SO in this case, we can change our topic modelling by time.



[Blei, Lafferty "Dynamic Topic Models ", https://mimno.infosci.cornell.edu/info6150/readings/dynamic_topic_models.pdf]

Summary

- Many topics are interpretable
- Works well with rare words
- Fast even for huge text collections
- Multicore & distributed implementations
- Many features can be added with extensions