Designing New Machine Learning Techniques

Presentation 1

Elijahu Ben-Michael   Mentor: Dr. Kevin Chen

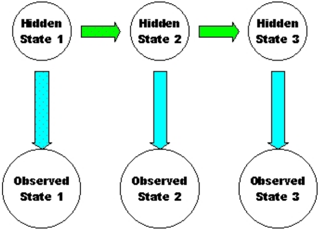
June 6, 2014

Outline of Talk

I **Hidden Markov Models and Chromatin Marks**

II Tensors and the Tensor Decomposition Problem

Hidden Markov Models



- Set of observations and hidden states
- Transition from hidden state to hidden state is a **Markov Process**
- Markov Condition: Given a set of random variables $\{H_t | t \in \mathbb{N}\}$, $P(H_{t+1} = h_{t+1} | H_t = h_t, \ldots, H_0 = h_0) = P(H_{t+1} = h_{t+1} | H_t = h_t)$

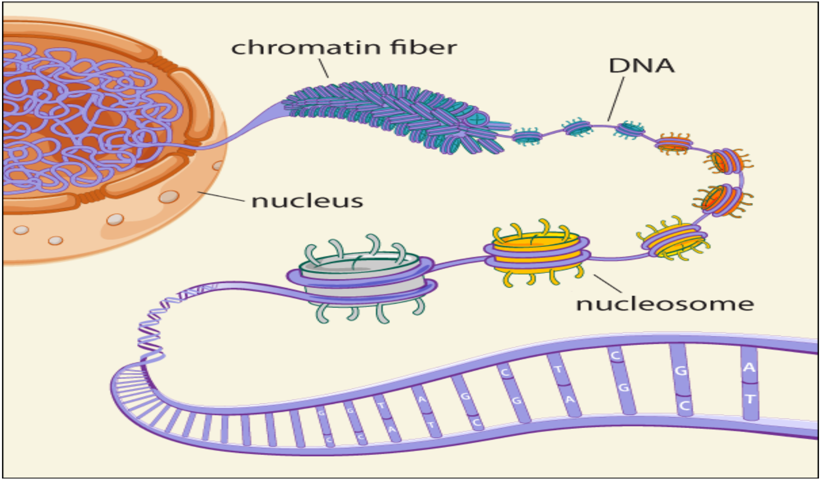
Chromatin

- Chromatin inside of nucleus hold genetic information
- Hundreds of slight modifications ("marks")

Question

Do chromatin marks correlate with functional elements in genome? If so, how?

Chromatin



Hidden Markov Models and Chromatin Marks

- Can use Hidden Markov Models to find the number of chromatin states and see what they do
- Observations: Chromatin Marks
- Hidden States: Chromatin States (associated with functional elements)

We can find the parameters of the Hidden Markov Model using a new technique

Outline of Talk

I Latent Variable Models and Chromatin Marks

II **Tensors and the Tensor Decomposition Problem**

Tensors

Tensor

A **Real pth-order Tensor** $T \in \otimes^p \mathbb{R}^n$ is a p-way array where $T_{i_1, i_2, \ldots, i_p}$ is the element at position $i_1, i_2, \ldots, i_p$

Tensor product and Tensor power

For any vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$, $\vec{v} \otimes \vec{w}$ denotes the **Tensor Product** of $\vec{v}$ and $\vec{w}$, and $\vec{v}^{\otimes p} = \vec{v} \otimes \vec{v} \otimes \vec{v} \otimes \ldots \otimes \vec{v} \in \otimes^p \mathbb{R}^n$ denotes the p-th **Tensor Power**

Example tensor product

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \vec{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}, \vec{v} \otimes \vec{w} = \begin{pmatrix} v_1 w_1 & v_1 w_2 \\ v_2 w_1 & v_2 w_2 \end{pmatrix}$$

Tensor Spectral Decomposition Problem

The problem

Given a 3rd order tensor $T \in \otimes^3 \mathbb{R}^n$ that is of the form $T = \sum_{j=1}^{n} \lambda_j \vec{v}_j^{\otimes 3}$ where $\{\vec{v}\}$ is an orthonormal basis of $\mathbb{R}^n$ and $\lambda_i > 0 \forall i$. Can we (approximately) find all of the $\vec{v}_i, \lambda_i$ pairs?

Questions

1. Are the $\vec{v}_i, \lambda_i$ pairs uniquely determined?
2. Are there any efficient algorithms to find them?

Matrix Analogy: EigenDecomposition

The problem

Given a matrix M of the form $M = \lambda_j \vec{v}\vec{v}^T$ where $\{\vec{v}\}$ is an orthonormal basis of $\mathbb{R}^n$ and $\lambda_i > 0 \forall i$. Can we (approximately) find all of the $\vec{v}_i, \lambda_i$ pairs?

Questions

1. Are the $\vec{v}_i, \lambda_i$ pairs unique? **YES**, if the eigenvalues are distinct
2. Are there any efficient algorithms to find them? **YES**

Proposition

The set $\{\vec{v}_i | \forall i\}$ *is the set of isolated local maximizers of* $f_t(\vec{x}) = \sum_{i,j,k} T_{i,j,k} x_i x_j x_k$

This tells us that we can identify the $\vec{v}_i, \lambda_i$ pairs

Tensor Power Method

A useful quadratic operator

For any real 3rd order Tensor T an $\vec{x} \in \mathbb{R}^n$, define $\phi_T(\vec{x}) = \sum_{i,j,k} T_{i,j,k} x_j x_k \vec{e}_i$, where $\vec{e}_i$ is the ith basis vector

For the required $T = \sum_{j=1}^{n} \lambda_j \vec{v}_j^{\otimes 3}$, $\phi_T(\vec{x}) = \sum_{i=1}^{n} \lambda_i (\vec{v}_i^T \vec{x})^2 \vec{v}_i$
Note: $\phi_T(\vec{v}_i) = \lambda_i \vec{v}_i$

Tensor Power Method

Power Iteration

Start with some $\vec{x}^{(0)}$ and for $k = 1, 2, 3, \ldots$
$\vec{x}^{(j)} = \phi_T(\vec{x}^{(j-1)})$

Proposition

For almost all initial $\vec{x}^{(0)}$, the sequence $\frac{\vec{x}^{(j)}}{||\vec{x}^{(j)}||}$ converges quadratically fast to a $\vec{v}_i$

Algorithm

- Pick an initial $\vec{x}^{(0)}$
- Run Power Iteration and return the approximations $(\hat{\vec{v}}_i, \hat{\lambda}_i)$
- Replace T with $T - \hat{\lambda}_i \hat{\vec{v}}_i^{\otimes 3}$ and repeat

Latent Variable Models

Many Latent Variable Models have tensor interpretations that are of the form $T = \sum_{j=1}^{n} \lambda_j \vec{v}_j^{\otimes 3}$ in which the parameters can be found from the eigenvectors and eigenvalues

- Exchangeable Topic Models
- Mixture of Gaussians
- Independent Component Analysis
- Latent Dirichlet Allocation
- Multi-view Models
- Hidden Markov Models

References

Anandkumar, A., Ge, R. Hsu, D., Kakade, S. and Telgarsky, M. *Tensor Decompositions for Learning Latent Variable Models* `arXiv:1210.7559 [cs.LG]` 2014

Baker, M. *Making sense of chromatin states. Nature Methods* 2011