Basic Assumptions for Efficient Model Representation

Michael Gutmann

Probabilistic Modelling and Reasoning (INFR11134)
School of Informatics, University of Edinburgh

Spring semester 2018

Recap

$$p(\mathbf{x}|\mathbf{y}_o) = \frac{\sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{y}_o, \mathbf{z})}{\sum_{\mathbf{x},\mathbf{z}} p(\mathbf{x}, \mathbf{y}_o, \mathbf{z})}$$

Assume that $\mathbf{x}, \mathbf{y}, \mathbf{z}$ each are $d = 500$ dimensional, and that each element of the vectors can take $K = 10$ values.

- Issue 1: To specify $p(\mathbf{x}, \mathbf{y}, \mathbf{z})$, we need to specify $K^{3d} - 1 = 10^{1500} - 1$ non-negative numbers, which is impossible.

  Topic 1: Representation What reasonably weak assumptions can we make to efficiently represent $p(\mathbf{x}, \mathbf{y}, \mathbf{z})$?

- Consider two assumptions
  1. only a limited number of variables may directly interact with each other (independence assumptions)
  2. the form of interaction is limited (often: parametric family assumptions)

  They can be used together or separately.

Program

1. Independence assumptions

2. Assumptions on form of interaction

Program

1. Independence assumptions
   - Definition and properties of statistical independence
   - Factorisation of the pdf and reduction in the number of directly interacting variables
2. Assumptions on form of interaction

Statistical independence

- Let $\mathbf{x}$ and $\mathbf{y}$ be two disjoint subsets of random variables. Then $\mathbf{x}$ and $\mathbf{y}$ are independent of each other if and only if (iff)

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$$

  for all possible values of $\mathbf{x}$ and $\mathbf{y}$; otherwise they are said to be dependent.

- We say that the joint *factorises* into a product of $p(\mathbf{x})$ and $p(\mathbf{y})$.
- Equivalent definition by the product rule (or by definition of conditional probability)

$$p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x})$$

  and all values of $\mathbf{x}$ and $\mathbf{y}$ where $p(\mathbf{y}) > 0$.

- Notation: $\mathbf{x} \perp\!\!\!\perp \mathbf{y}$
- Variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent iff

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \prod_{i=1}^{n} p(\mathbf{x}_i)$$

Conditional statistical independence

- The characterisation of statistical independence extends to conditional pdfs (pmfs) $p(\mathbf{x}, \mathbf{y}|\mathbf{z})$.
- The condition $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$ becomes $p(\mathbf{x}, \mathbf{y}|\mathbf{z}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{y}|\mathbf{z})$
- The equivalent condition $p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x})$ becomes $p(\mathbf{x}|\mathbf{y}, \mathbf{z}) = p(\mathbf{x}|\mathbf{z})$
- We say that $\mathbf{x}$ and $\mathbf{y}$ are conditionally independent given $\mathbf{z}$ iff, for all possible values of $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ with $p(\mathbf{z}) > 0$:

$$p(\mathbf{x}, \mathbf{y}|\mathbf{z}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{y}|\mathbf{z}) \quad \text{or}$$

$$p(\mathbf{x}|\mathbf{y}, \mathbf{z}) = p(\mathbf{x}|\mathbf{z}) \quad (\text{for } p(\mathbf{y}, \mathbf{z}) > 0)$$

- Notation: $\mathbf{x} \perp\!\!\!\perp \mathbf{y} \mid \mathbf{z}$

The impact of independence assumptions

- The key is that the independence assumption leads to a partial factorisation of the pdf (pmf).
- For example, if $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are independent of each other, then

$$p(\mathbf{x}, \mathbf{y}, \mathbf{z}) = p(\mathbf{x})p(\mathbf{y})p(\mathbf{z})$$

- If $\dim(\mathbf{x}) = \dim(\mathbf{y}) = \dim(\mathbf{z}) = d$, and each element of the vectors can take K values, factorisation reduces the numbers that need to be specified ("parameters") from $K^{3d} - 1$ to $3(K^d - 1)$.
- If all variables were independent: $3d(K - 1)$ numbers needed.

  For example: $10^{1500} - 1$ vs. $3(10^{500} - 1)$ vs $1500(10 - 1) = 13500$

- But full independence (factorisation) assumption is often too strong and does not hold.

The impact of independence assumptions

- Conditional independence assumptions are a powerful middle-ground.
- For $p(\mathbf{x}) = p(x_1, \ldots, x_d)$, we have by the product rule:

$$p(\mathbf{x}) = p(x_d | x_1, \ldots x_{d-1}) p(x_1, \ldots, x_{d-1})$$

- If, for example, $x_d \perp\!\!\!\perp x_1, \ldots, x_{d-4} \,|\, x_{d-3}, x_{d-2}, x_{d-1}$, we have

$$p(x_d | x_1, \ldots, x_{d-1}) = p(x_d | x_{d-3}, x_{d-2}, x_{d-1})$$

- If the x_i can take K different values:

  $p(x_d | x_1, \ldots, x_{d-1})$ specified by $K^{d-1} \cdot (K-1)$ numbers

  $p(x_d | x_{d-3}, x_{d-2}, x_{d-1})$ specified by $K^3 \cdot (K-1)$ numbers

  For $d = 500, K = 10$: $10^{499} \cdot 9 \approx 10^{500}$ vs $9000 \approx 10^4$.

Program

1. Independence assumptions
2. Assumptions on form of interaction
   - Parametric model to restrict how a given number of variables may interact

Assumption 2: limiting the form of the interaction

- The (conditional) independence assumption limits the number of variables that may directly interact with each other, e.g. x_d only directly interacted with $x_{d-3}, x_{d-2}, x_{d-1}$.
- How x_d interacts with the three variables, however, was not restricted.
- Assumption 2: We restrict how a given number of variables may interact with each other.
- For example, for $x_i \in \{0, 1\}$, we may assume that $p(x_d|x_1, \ldots, x_{d-1})$ is specified as

$$p(x_d = 1|x_1, \ldots, x_{d-1}) = \frac{1}{1 + \exp\left(-w_0 - \sum_{i=1}^{d-1} w_i x_i\right)}$$

  with d free numbers ("parameters") $w_0, \ldots, w_{d-1}$.
- d vs 2^{d-1} numbers

Program recap

We asked: What reasonably weak assumptions can we make to efficiently represent a probabilistic model?

1. Independence assumptions
   - Definition and properties of statistical independence
   - Factorisation of the pdf and reduction in the number of directly interacting variables

2. Assumptions on form of interaction
   - Parametric model to restrict how a given number of variables may interact