Generative Adversarial Networks

Machine Learning and Data Mining

Maxim Borisyak

National Research University Higher School of Economics (HSE)

November 21, 2017

Generative models

Generative models

- Informally, given samples we wish to learn underlying distribution in form of sampling procedure.
- Formally, given samples of a random variable X, we wish to find X', so that:

$$P(X) \approx P(X')$$

Types of generative models

Fitting density function $P(X)$ (or a function $f(x)$ proportional to the density):

- partition function Z for $f(x)$ might be an issue:
  - for energy models Contrastive Devergence keeps Z finite (ideally constant);
- sampling might be an issue:
  - Gibbs sampling works ok if $P(x^i \mid x^{-i})$ is analytically known and simple enough.

Going deep:

- RBM is intrinsically one-layer model;
- Deep Boltzmann machines:
  - Gibbs sampling becomes less efficient than for RBM.

Types of generative procedures

Options for defining a random variables:

- specify $P(X)$ and use general sampling algorithm (e.g. Gibbs sampling);
- **learn sampling procedure directly**, e.g.:

$$\begin{aligned} X &= f(Z); \\ Z &\sim \text{SimpleDistribution}; \end{aligned}$$

Fitting Distributions

Notation: Q - ground truth distribution, P - model.

Maximum Likelihood:

$$\begin{aligned}
\mathcal{L} &= \sum_i \log P(x_i) \approx \mathop{\mathbb{E}}_{x \sim Q} \log P(x) \to_P \min; \\
\mathrm{KL}(Q \,\|\, P) &= \mathop{\mathbb{E}}_{x \sim Q} \log Q(x) - \mathop{\mathbb{E}}_{x \sim Q} \log P(x) \to_P \min\,.
\end{aligned}$$

Jenson-Shenon:

$$\begin{aligned}
\mathrm{JS}(P, Q) &= \frac{1}{2}\left[\mathrm{KL}(P \,\|\, M) + \mathrm{KL}(Q \,\|\, M)\right] \to_P \min; \\
M &= \frac{1}{2}(P + Q).
\end{aligned}$$

Approximating JS distance

$$\mathrm{JS}(P,Q) = \frac{1}{2}\left[\mathop{\mathbb{E}}_{x\sim P}\log\frac{P(x)}{M(x)} + \mathop{\mathbb{E}}_{x\sim Q}\log\frac{Q(x)}{M(x)}\right] =$$

$$\frac{1}{2}\left[\mathop{\mathbb{E}}_{x\sim P}\log\frac{P(x)}{P(x)+Q(x)} + \mathop{\mathbb{E}}_{x\sim Q}\log\frac{Q(x)}{P(x)+Q(x)}\right] + \log 2 =$$

$$\mathop{\mathbb{E}}_{x\sim M}\frac{P(x)}{P(x)+Q(x)}\log\frac{P(x)}{P(x)+Q(x)} +$$

$$\mathop{\mathbb{E}}_{x\sim M}\frac{Q(x)}{P(x)+Q(x)}\log\frac{Q(x)}{P(x)+Q(x)} + \log 2$$

Approximating JS distance

$$\mathrm{JS}(P,Q) = \mathop{\mathbb{E}}_{x\sim M} \frac{P(x)}{P(x)+Q(x)} \log \frac{P(x)}{P(x)+Q(x)} + \mathop{\mathbb{E}}_{x\sim M} \frac{Q(x)}{P(x)+Q(x)} \log \frac{Q(x)}{P(x)+Q(x)} + \log 2$$

Let's introduce y: $y=1$ if x is sampled from P and $y=0$ for Q:

$$\mathrm{JS}(P,Q) - \log 2 =$$

$$\mathop{\mathbb{E}}_{x\sim M} \left[P(y=1\mid x)\log P(y=1\mid x) + P(y=0\mid x)\log P(y=0\mid x)\right] =$$

$$\mathop{\mathbb{E}}_{x\sim M} \left[\mathop{\mathbb{E}}_{y}[y\mid x]\log P(y=1\mid x) + (1-\mathop{\mathbb{E}}_{y}[y\mid x])\log P(y=0\mid x)\right]$$

Approximating JS distance

$$\mathrm{JS}(P, Q) - \log 2 =$$

$$\mathop{\mathbb{E}}_{x \sim M} \left[\mathop{\mathbb{E}}_{y} [y \mid x] \log P(y = 1 \mid x) + (1 - \mathop{\mathbb{E}}_{y} [y \mid x]) \log P(y = 0 \mid x) \right] =$$

$$\mathop{\mathbb{E}}_{x \sim M, y} y \log P(y = 1 \mid x) + (1 - y) \log P(y = 0 \mid x)$$

$$\boxed{\mathrm{JS}(P, Q) = \log 2 - \min_{f} \left[\text{cross–entropy}(f \parallel P, Q) \right]}$$