Deep Learning - Foundations and Concepts

Chapter 3. Standard Distributions

nonlineark@github

February 5, 2025

Outline

Bernoulli distribution

- Consider a binary random variable $x \in \{0, 1\}$ and a parameter $0 \leq \mu \leq 1$, such that $p(x = 1) = \mu$ and $p(x = 0) = 1 - \mu$.
- Probability distribution: $\mathrm{Bern}(x; \mu) = \mu^x (1 - \mu)^{1-x}$.
- Expectation: $E(x) = \mu$.
- Variance: $\mathrm{var}(x) = \mu(1 - \mu)$.

Bernoulli distribution

Model the Bernoulli distribution given observations $\{x_1, \dots, x_N\}$.

$$
\begin{aligned}
p(x_1, \dots, x_N; \mu) &= \prod_{n=1}^{N} \mu^{x_n}(1-\mu)^{1-x_n} \\
\log p(x_1, \dots, x_N; \mu) &= \sum_{n=1}^{N} (x_n \log \mu + (1-x_n)\log(1-\mu)) \\
&= \log \mu \sum_{n=1}^{N} x_n + \log(1-\mu)(N - \sum_{n=1}^{N} x_n) \\
\mu_{ML} &= \frac{1}{N}\sum_{n=1}^{N} x_n
\end{aligned}
$$

Binomial distribution

- Consider a random variable $m = \sum_{n=1}^{N} x_n$, where x_n are independent random variables obey Bernoulli distribution with parameter μ.
- Probability distribution: $\mathrm{Bin}(m; N, \mu) = \binom{N}{m} \mu^m (1-\mu)^{N-m}$.
- Expectation: $E(m) = N\mu$.
- Variance: $\mathrm{var}(m) = N\mu(1-\mu)$.

Multinomial distribution

- Consider a random variable $x \in \{\mathrm{e}_1, \dots, \mathrm{e}_K\}$ and a parameter $\mu \in \mathbb{R}^K$, such that $p(x = \mathrm{e}_k) = \mu_k$.
- Probability distribution: $p(x; \mu) = \prod_{k=1}^{K} \mu_k^{x_k}$.
- Expectation: $E(x) = \mu$.
- Covariance: $\mathrm{cov}(x) = \mathrm{diag}(\mu_1(1 - \mu_1), \dots, \mu_K(1 - \mu_K))$.

Multinomial distribution

Model the generalized Bernoulli distribution given observations $x^1, \ldots, x^N$.

$$p(x^1, \ldots, x^N; \mu) = \prod_{n=1}^{N} \prod_{k=1}^{K} \mu_k^{x_k^n}$$

$$\log p(x^1, \ldots, x^N; \mu) = \sum_{n=1}^{N} \sum_{k=1}^{K} x_k^n \log \mu_k = \sum_{k=1}^{K} (\sum_{n=1}^{N} x_k^n) \log \mu_k$$

$$\mu_{ML} = \frac{1}{N} \sum_{n=1}^{N} x^n$$

For the last step, we used Lagrange multiplier to take into the constraint $\sum_{k=1}^{K} \mu_k = 1$.

Multinomial distribution

- Consider a random variable $m = \sum_{n=1}^{N} x^n$, where x^n are independent random variables obey the generalized Bernoulli distribution with parameter μ.
- Probability distribution: $\mathrm{Mult}(m; N, \mu) = \frac{N!}{\prod_{k=1}^{K} m_k!} \prod_{k=1}^{K} \mu_k^{m_k}$.
- Expectation: $E(m) = N\mu$.
- Covariance: $\mathrm{cov}(m) = N(\mathrm{diag}(\mu_1, \ldots, \mu_K) - \mu\mu^T)$.