Machine Learning 2 - Cheat Sheet
Andrea Jemmett - April 16, 2016

1 Distributions

Binary	Bernoulli	Binomial	Beta
Discrete	Categorical	Multinomial	Dirichlet

1.1 Bernoulli

$\text{Ber}(x|n) = \mu^x(1-\mu)^{1-x}$, $\quad \mathbb{E}[x] = \mu$, $\quad \text{V}ar[x] = \mu - \mu^2$,
$P(D,\mu) = \prod_{n=1}^N \mu^{x_n}(1-\mu)^{1-x_n}$, $\quad \mu_{\text{ML}} = \frac{1}{N}\sum_{n=1}^N x_n$

1.2 Binomial

$\text{Bin}(m|N,\mu) = \binom{N}{m}\mu^m(1-\mu)^{N-m}$, $\quad \frac{n!}{k!(n-k)!} = \binom{n}{k}$,
$\mathbb{E}[m] = N\mu$, $\quad \text{V}ar[m] = N\mu(1-\mu)$, $\quad \mu_{\text{ML}} = \frac{m}{N}$

1.3 Categorical

$p(\mathbf{x}|\boldsymbol{\mu}) = \prod_k \mu_k^{x_k}$, $\quad \boldsymbol{\mu} \in \{0,1\}^K$, $\quad \sum_k \mu_k = 1$, $\quad \boldsymbol{\mu}_{\text{ML}} = \frac{\mathbf{m}}{N}$,
$m_k = \sum_n x_{nk}$, $\quad \text{Mult}(m_1 \ldots, m_k|N,\boldsymbol{\mu}) = (\frac{N!}{m_1!,\ldots,m_k}\prod_k)\mu_k^{mk}$

1.4 Beta

$\text{Beta}(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}$, [1] $\quad \mathbb{E}[\mu] = \frac{a}{a+b}$,
$\text{V}ar[x] = \frac{ab}{(a+b)^2(a+b+1)}$, $\quad p(\mu|m,l,a,b) \propto \mu^{m+a-1}(1-\mu)^{l+b-1}$

1.5 Gamma Distribution

$\text{Gamma}(\tau|a,b) = \frac{b^a}{\Gamma(a)}\tau^{a-1}e^{-b\tau}$, $\quad \mathbb{E}[\tau] = \frac{a}{b}$, $\quad \text{V}ar[\tau] = \frac{a}{b^2}$,
$\text{mode}[\tau] = \frac{a-1}{b}$ for $a \geq 1$, $\quad \mathbb{E}[\ln\tau] = \psi(a) - \ln b$,
$H(\tau) = ln\Gamma(a) - (a-1)\psi(a) - \ln b + a$

1.6 Multinomial Distributions

$\mathbf{x} = \{0,0,0,0,1,0,0\}^T$, $\sum_{k=1}^K x_k = 1$, $p(x|\mu) = \prod_{k=1}^K \mu_k^{x_k}$, $\sum_{k=1}^K \mu_k = 1$,
$\mu_{ML} = \frac{m}{N}$, $m_k = \sum_{k=1}^K x_{nk}$

1.7 Dirichlet

$\text{Dir}(\mu|\alpha) = \frac{\Gamma(\sum_k a_k)}{\Gamma a_1 \ldots \Gamma a_k}\prod_{k=1}^K \mu_k^{a_k - 1}$

1.8 Gaussian

$\mathcal{N}(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{1}{2\sigma^2}(x-\mu)^2)$,
$\mathcal{N}(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}}\frac{1}{\sqrt{|\Sigma|}}\exp(-\frac{1}{2}(x-\mu)^T)\Sigma^{-1}(x-\mu))$

1.9 ML for the Gaussian

$\ln p(X|\mu,\Sigma) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^N(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)$,
$\mu_{ML} = 1/N\sum_{n=1}^N x_n$, $\Sigma_{ML} = 1/N\sum_{n=1}^N(x_n-\mu)^T(x_n-\mu)$

1.10 Stochastic gradient descent Gaussian

$\max\ P(x_1,\cdots,x_n|\theta)$, $\theta^N = \theta^{N-1} + \alpha_{N-1}\frac{\partial}{\partial\theta^{N-1}}\ln\ p(x_n|\theta^{N-1})$

1.11 Marginal and Conditional Gaussians

Given $p(x) = \mathcal{N}(x|\mu,\Lambda^{-1})$ and $p(y|x) = \mathcal{N}(y|Ax+b,L^{-1})$. We get $p(y) = \mathcal{N}(y|A\mu+b,L^{-1}+A\Lambda^{-1}A^T)$ and $p(x|y) = \mathcal{N}(x|\Sigma\{A^TL(y-b)+\Lambda\mu\},\Sigma)$, where $\Sigma = (\Lambda + A^TLA)^{-1}$.

1.12 Student's T distribution

The heavy tail of the student-t distribution makes it more robust against outliers.
$St(x|\mu,\lambda,\nu) = \frac{\Gamma(\nu/2+1/2)}{\Gamma(\nu/2)}(\frac{\lambda^{1/2}}{(\pi\nu)^{D/2}})[1+\frac{\lambda(x-\mu)^2}{\nu}]^{-\nu/2-D/2}$,
$f_x(x) = \frac{\Gamma[(\nu+p)/2]}{\Gamma(\nu/2)\nu^{p/2}\pi^{p/2}|\Sigma|^{1/2}[1+1/\nu(x-\mu)^T\Sigma^{-1}(x-\mu)]^{(\nu+p)/2}}$
$\mathbb{E}(\mathbf{x}) = \frac{\Gamma(D/2+v/2)}{\Gamma(v/2)}\frac{|\Lambda|^{1/2}}{(\pi v)^{D/2}}\int[1+\frac{(x-\mu)^T\Lambda(x-\mu)}{v}]^{-D/2-v/2}x\mathrm{d}x$

---

[1] $\Gamma(x) = \int_0^1 u^{x-1}e^{-u} = 1$, $\quad \Gamma(x+1) = \Gamma(x)x$, $\quad \Gamma(x+1) = x!$