

My solutions to
Deep Learning: Foundations and Concepts

Dario Miro Konopatzki

3 Standard Distributions

3.1

$$\begin{aligned}\sum p_{X;\mu} &= \sum_{x \in \{0,1\}} \mu^x (1-\mu)^{1-x} \\ &= \mu^0 (1-\mu)^{(1-0)} + \mu^1 (1-\mu)^{1-1} \\ &= 1 \cdot (1-\mu) + \mu \cdot 1 \\ &= 1 - \mu + \mu \\ &= 1\end{aligned}$$

$$\begin{aligned}\mathbb{E}[X] &= \sum_{x \in \{0,1\}} x \mu^x (1-\mu)^{1-x} \\ &= 0 + 1 \cdot \mu^1 (1-\mu)^{1-1} \\ &= \mu\end{aligned}$$

$$\begin{aligned}\mathbb{V}[X] &= \mathbb{E}[x^2] - \mathbb{E}[x]^2 \\ &= \sum_{x \in \{0,1\}} x^2 \mu^x (1-\mu)^{1-x} - \mu^2 \\ &= 0 + 1^2 \cdot \mu^1 (1-\mu)^{1-1} - \mu^2 \\ &= \mu - \mu^2 \\ &= \mu(1-\mu)\end{aligned}$$

$$\begin{aligned}\mathbb{H}[x] &= \mathbb{E}[-\log_2 p] \\ &= \sum_{x \in \{0,1\}} -p(x) \log_2 p(x)\end{aligned}$$

$$\begin{aligned}
&= \sum_{x \in \{0,1\}} -\mu^x (1-\mu)^{1-x} \log_2 (\mu^x (1-\mu)^{1-x}) \\
&= \sum_{x \in \{0,1\}} -\mu^x (1-\mu)^{1-x} (x \log_2 \mu + (1-x) \log_2 (1-\mu)) \\
&= -\mu^0 (1-\mu)^{1-0} (0 \cdot \log_2 \mu + (1-0) \log_2 (1-\mu)) \\
&\quad - \mu^1 (1-\mu)^{1-1} (1 \cdot \log_2 \mu + (1-1) \log_2 (1-\mu)) \\
&= (1-\mu) \log_2 (1-\mu) - \mu \log_2 \mu
\end{aligned}$$

3.2

$$\begin{aligned}
\sum p_{X;\mu} &= \sum_{\{-1,1\}} \left(\frac{1-\mu}{2} \right)^{\frac{1-x}{2}} \left(\frac{1+\mu}{2} \right)^{\frac{1+x}{2}} \\
&= \left(\frac{1-\mu}{2} \right)^{\frac{1-(-1)}{2}} \left(\frac{1+\mu}{2} \right)^{\frac{1+(-1)}{2}} + \left(\frac{1-\mu}{2} \right)^{\frac{1-1}{2}} \left(\frac{1+\mu}{2} \right)^{\frac{1+1}{2}} \\
&= \frac{1-\mu}{2} \cdot 1 + 1 \cdot \frac{1+\mu}{2} \\
&= \frac{1-\mu+1+\mu}{2} \\
&= 1
\end{aligned}$$

Mean

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{\{-1,1\}} x \left(\frac{1-\mu}{2} \right)^{\frac{1-x}{2}} \left(\frac{1+\mu}{2} \right)^{\frac{1+x}{2}} \\
&= -1 \cdot \frac{1-\mu}{2} + 1 \cdot \frac{1+\mu}{2} \\
&= \frac{-1+\mu+1+\mu}{2} \\
&= \mu
\end{aligned}$$

Variance

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{\{-1,1\}} x^2 \left(\frac{1-\mu}{2} \right)^{\frac{1-x}{2}} \left(\frac{1+\mu}{2} \right)^{\frac{1+x}{2}} \\
&= 1 \cdot \frac{1-\mu}{2} + 1 \cdot \frac{1+\mu}{2} \\
&= 1
\end{aligned}$$

$$\begin{aligned}\mathbb{V}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= 1 - \mu^2\end{aligned}$$

Entropy

$$\begin{aligned}\mathbb{H}[X] &= \mathbb{E}\left[\log \frac{1}{p_{X;\mu}}\right] \\ &= \sum_{x \in \{-1,1\}} \left(\log \frac{1}{p_{X;\mu}(x)}\right) p_{X;\mu}(x) \\ &= \left(\log \frac{2}{1-\mu}\right) \frac{1-\mu}{2} + \left(\log \frac{2}{1+\mu}\right) \frac{1+\mu}{2} \\ &= (\log(2) - \log(1-\mu)) \frac{1-\mu}{2} + (\log(2) - \log(1+\mu)) \frac{1+\mu}{2} \\ &= \log(2) \frac{1-\mu+1+\mu}{2} - \log(1-\mu) \frac{1-\mu}{2} - \log(1+\mu) \frac{1+\mu}{2} \\ &= \log(2) - \log(1-\mu) \frac{1-\mu}{2} - \log(1+\mu) \frac{1+\mu}{2}\end{aligned}$$

3.3

$$\begin{aligned}\binom{N}{m} + \binom{N}{m-1} &= \frac{N!}{(N-m)!m!} + \frac{N!}{(N-(m-1))!(m-1)!} \\ &= \frac{N!(N-(m-1))!(m-1)! + N!(N-m)!m!}{(N-m)!m!(N-(m-1))!(m-1)!} \\ &= \frac{N!((N-m+1)!(m-1)! + (N-m)!m!)}{(N-m)!m!(N+1-m))!(m-1)!} \\ &= \frac{N!(\cancel{(N-m)!}(\cancel{m-1})!((N-m+1)+m))}{(\cancel{(N-m)!}m!(N+1-m))!(\cancel{m-1})!} \\ &= \frac{N!(N+1)}{(N+1-m)!m!} \\ &= \frac{(N+1)!}{((N+1)-m)!m!} \\ &= \binom{N+1}{m}\end{aligned}$$

$$\mathbf{N} = \mathbf{0}$$

$$\begin{aligned}\sum_{m=0}^0 \binom{0}{m} x^m &= \binom{0}{0} x^0 \\ &= \frac{0!}{(0-0)!0!} \cdot 1 \\ &= 1\end{aligned}$$

$$\mathbf{N} \rightarrow \mathbf{N} + \mathbf{1}$$

$$\begin{aligned}(1+x)^{N+1} &= (1+x)(1+x)^N \\ &= (1+x) \sum_{m=0}^N \binom{N}{m} x^m \quad \text{induction hypothesis} \\ &= \sum_{m=0}^N \binom{N}{m} x^m + \sum_{m=0}^N \binom{N}{m} x^{m+1} \\ &= \binom{N}{0} x^0 + \sum_{m=1}^N \binom{N}{m} x^m \\ &\quad + \sum_{m=0}^{N-1} \binom{N}{m} x^{m+1} + \binom{N}{N} x^{N+1} \\ &= 1 \cdot x^0 + \sum_{m=1}^N \binom{N}{m} x^m \\ &\quad + \sum_{m=1}^N \binom{N}{m-1} x^m + 1 \cdot x^{N+1} \\ &= 1 \cdot x^0 + \sum_{m=1}^N \left(\binom{N}{m} + \binom{N}{m-1} \right) x^m + 1 \cdot x^{N+1} \\ &= \binom{N+1}{0} x^0 + \sum_{m=1}^N \binom{N+1}{m} x^m \\ &\quad + \binom{N+1}{N+1} x^{N+1} \\ &= \sum_{m=0}^{N+1} \binom{N+1}{m} x^m\end{aligned}$$

Normalization

$$\begin{aligned}
\sum_{m=0}^N \binom{N}{m} \mu^m (1-\mu)^{N-m} &= (1-\mu)^N \sum_{m=0}^N \binom{N}{m} (\mu(1-\mu)^{-1})^m \\
&= (1-\mu)^N (1 + \mu(1-\mu)^{-1})^N \quad \text{binom. thm.} \\
&= ((1-\mu)(1 + \mu(1-\mu)^{-1}))^N \\
&= ((1-\mu) + (1-\mu)\mu(1-\mu)^{-1})^N \\
&= (1-\mu + \mu)^N \\
&= 1
\end{aligned}$$

3.4

Expectation

For $\mu = 0$: $\mathbb{E}[M] = \sum_{m=0}^N m \binom{N}{m} 0^m (1-0)^{N-m} = 0 = N \cdot 0$. For $\mu \neq 0$:

$$\begin{aligned}
0 &= \frac{\partial}{\partial \mu} 1 \\
&= \frac{\partial}{\partial \mu} \overbrace{\sum_{m=0}^N \binom{N}{m} \mu^m (1-\mu)^{N-m}}^{=1 \text{ by 3.3}} \\
&= \sum_{m=0}^N \binom{N}{m} (m\mu^{m-1}(1-\mu)^{N-m} + \mu^m(N-m)(1-\mu)^{N-m-1}(-1)) \\
&= \sum_{m=0}^N \binom{N}{m} (m\mu^m(1-\mu)^{N-m}(\mu^{-1} + (1-\mu)^{-1}) - N\mu^m(1-\mu)^{N-m-1}) \\
&\stackrel{(\star)}{\iff} \sum_{m=0}^N \binom{N}{m} m\mu^m(1-\mu)^{N-m} = \sum_{m=0}^N \binom{N}{m} N\mu^m(1-\mu)^{N-m-1}\mu(1-\mu) \\
&\iff \sum_{m=0}^N m \binom{N}{m} \mu^m(1-\mu)^{N-m} = N\mu \underbrace{\sum_{m=0}^N \binom{N}{m} \mu^m(1-\mu)^{N-m}}_{=1 \text{ by 3.3}} \\
&\iff \mathbb{E}[M] = N\mu
\end{aligned}$$

$$(\star) \quad \mu^{-1} + (1-\mu)^{-1} = \mu^{-1}(1-\mu)^{-1}((1-\mu) + \mu) = \mu^{-1}(1-\mu)^{-1}$$

Variance

For $\mu = 0$: $\mathbb{V}[M] = \sum_{m=0}^N \overbrace{(m - \mathbb{E}[M])^2}^{=m-N \cdot 0=m} \binom{N}{m} 0^m (1-0)^{N-m} = 0 = N \cdot 0(1-0)$.

For $\mu \neq 0$:

$$\begin{aligned}
N &= \frac{\partial}{\partial \mu} (N\mu) \\
&= \frac{\partial}{\partial \mu} \mathbb{E}[M] \\
&= \frac{\partial}{\partial \mu} \sum_{m=0}^N m \binom{N}{m} \mu^m (1-\mu)^{N-m} \\
&= \sum_{m=0}^N m \binom{N}{m} (m\mu^m (1-\mu)^{N-m} (\mu^{-1} + (1-\mu)^{-1}) \\
&\quad - N\mu^m (1-\mu)^{N-m-1})
\end{aligned}$$

$\xLeftrightarrow{(*)}$

$$\begin{aligned}
\sum_{m=0}^N m^2 \binom{N}{m} \mu^m (1-\mu)^{N-m} &= N\mu(1-\mu) + N\mu \sum_{m=0}^N m \binom{N}{m} \mu^m (1-\mu)^{N-m} \\
&\iff \mathbb{E}[M^2] = N\mu(1-\mu) + N\mu \mathbb{E}[M] \\
&= N\mu(1-\mu) + (N\mu)^2
\end{aligned}$$

and consequently $\mathbb{V}[M] = \mathbb{E}[M^2] - \mathbb{E}[M]^2 = N\mu(1-\mu) + (N\mu)^2 - (N\mu)^2 = N\mu(1-\mu)$.

3.5

We assume Σ invertible. By definition of multivariate normal Σ positive semi-definite, which combined with the former means Σ positive definite.

$$\frac{\partial}{\partial x} p(x) = \frac{\partial}{\partial x} \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{\sqrt{\det \Sigma}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)}$$

$$\begin{aligned}
&\stackrel{(\star)}{=} p(x) \left(-\frac{1}{2}(x - \mu)^\top \left(\Sigma^{-1} + (\Sigma^{-1})^\top \right) \right) \\
&\stackrel{(\dagger)}{=} -p(x)(x - \mu)^\top \Sigma^{-1}
\end{aligned}$$

By assumption Σ invertible, i.p. $\Sigma^{-1} \neq 0$. Since Σ positive definite, $\det \Sigma > 0$ and consequently $p(x) > 0$ f.a. x . Taken together $\frac{\partial}{\partial x} p(x) = 0$ only if $x = \mu$.

$$\begin{aligned}
\frac{\partial}{\partial x} \nabla p(x) &= - \left(p(x)\Sigma^{-1} + \Sigma^{-1}(x - \mu) \left(-p(x)(x - \mu)^\top \Sigma^{-1} \right) \right) \\
&= p(x)\Sigma^{-1} \left((x - \mu)(x - \mu)^\top \Sigma^{-1} - \mathbb{I} \right)
\end{aligned}$$

Which is $p(\mu)\Sigma^{-1}$ for $x = \mu$. Since Σ positive definite, so is Σ^{-1} , and with $p(\mu) > 0$ this gives $p(\mu)\Sigma$ positive definite. Hence $-p(\mu)\Sigma$ negative definite, and consequently μ maximum of $p(x)$, implying μ mode of multivariate Gaussian.

(\star) Matrix calculus; numerator layout

(\dagger) Σ symmetric & inverse of symmetric matrix symmetric