Contents

Chapter 1

Solutions for exercises to chapter 1

Problem 1.1 [closed form solution to polynomial regression]

We use a slightly better notation to write this problem. Let X be the matrix of the form

$$X = \begin{bmatrix} x_1^0 & x_1^1 & \cdots & x_1^M \\ x_2^0 & x_2^1 & \cdots & x_2^M \\ \vdots & \vdots & \ddots & \vdots \\ x_N^0 & x_N^1 & \cdots & x_N^M \end{bmatrix}, \quad t = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}$$

The the problem can be rewritten in the following form:

$$E(w) = \frac{1}{2}\left((Xw-t)^T(Xw-t)\right).$$

Now we differentiate w.r.t w, note that

$$\begin{aligned} E(w+h) &= \frac{1}{2}\left(X(w+h)-t\right)^T\left(X(w+h)-t\right) \\ &= \frac{1}{2}\left((Xw-t)^T+(Xh)^T\right)(Xw-t+Xh) \\ &= \frac{1}{2}\left[(Xw-t)^T(Xw-t)+(Xw-t)^T Xh+(Xh)^T(Xw-t)+(Xh)^T(Xh)\right] \\ &= E(w)+\left\langle (Xw-t)^T, Xh\right\rangle+\frac{1}{2}\langle Xh, Xh\rangle \\ &= E(w)+\left\langle X^T(Xw-t), h\right\rangle+\frac{1}{2}\langle Xh, Xh\rangle. \end{aligned}$$

Note that $\left\langle X^T(Xw-t), h\right\rangle \in \mathrm{Hom}(\mathbb{R}^{M+1}, \mathbb{R})$ and

$$\frac{1}{2}\langle Xh, Xh\rangle \le \frac{1}{2}\|Xh\|\,\|Xh\| \le \frac{C}{2}\|X\|_\infty^2\,\|h\| \xrightarrow{\|h\|\to 0} 0,$$

it follows that $\nabla E(w) = X^T(Xw-t)$. Set it to zero and we get

$$X^T(Xw-t)=0 \iff X^T Xw = X^\top t.$$

So X^TX is the A proposed in the problem.

$$\left[X^TX\right]_{ij} = \sum_{n=1}^{N}\left(x_n^i x_n^j\right) = \sum_{n=1}^{N} x_n^{i+j}, \text{ and } \left[X^Tt\right]_i = \sum_{n=1}^{N} x_n^i t_n,$$

as desired.

Problem 1.2 [closed form solution to regularized polynomial regression]

We use the same notation as in the previous problem and still rewrite the loss function in matrix form as follows:

$$\widetilde{E}(w) = \frac{1}{2}\left\langle Xw - t, Xw - t\right\rangle + \frac{\lambda}{2}\left\langle w, w\right\rangle.$$

Still we differentiate the expression. Note that if we let $\varphi(w) = \frac{\lambda}{2}\left\langle w, w\right\rangle$,we have that

$$\begin{aligned}
\varphi(w+h) &= \frac{\lambda}{2}\left(w+h\right)^T\left(w+h\right) \\
&= \frac{\lambda}{2}\left(w^Tw + w^Th + h^Tx + \|h\|\right) \\
&= \varphi\left(w\right) + \left\langle \lambda w, h\right\rangle + \underbrace{\frac{\lambda}{2}\|h\|}_{=o(\|h\|)}.
\end{aligned}$$

Therefore, $\nabla\varphi(w) = \lambda w$, and as a result

$$\nabla\widetilde{E}(w) = \nabla E(w) + \nabla\varphi(w) = X^T(Xw - t) + \lambda w.$$

Setting it to zero:

$$X^T(Xw - t) + \lambda w = 0 \iff (X^TX + \lambda I)w = X^Tt.$$

Hence, $(X^TX + \lambda I)$ and X^Tt are the corresponding matrices.

Problem 1.3 [bayes formula warm up]

According to the Bayes formula, we get that

$$\begin{aligned}
P\left(\text{apple}\right) &= P\left(\text{apple}|\text{r}\right)P\left(\text{r}\right) + P\left(\text{apple}|\text{g}\right)P\left(\text{g}\right) + P\left(\text{apple}|\text{b}\right)P\left(\text{b}\right) \\
&= \frac{3}{10}\cdot\frac{2}{10} + \frac{1}{2}\frac{2}{10} + \frac{3}{10}\frac{6}{10} = \frac{17}{50}.
\end{aligned}$$

And again, we can use formula to get

$$\begin{aligned}
P\left(\text{g}|\text{orange}\right) &= \frac{P\left(\text{orange}|\text{g}\right)P\left(\text{g}\right)}{P\left(\text{orange}|\text{g}\right)P\left(\text{g}\right) + P\left(\text{orange}|\text{b}\right)P\left(\text{b}\right) + P\left(\text{orange}|\text{r}\right)P\left(\text{r}\right)} \\
&= \frac{\frac{3}{10}\frac{6}{10}}{\frac{3}{10}\frac{6}{10} + \frac{2}{10}\frac{1}{2} + \frac{2}{10}\frac{4}{10}} \\
&= \frac{1}{2}.
\end{aligned}$$

Problem 1.4 [nonlinear transform of likelihood function doesn't preserve its extrema]

We first observe that if x_* maximizes the likelihood function $p_x(x)$, then $p'_x(x_*) = 0$. By chain rule, we have that

$$\begin{aligned}\frac{dp_x(g(y))\,|g'(y)|}{dy} &= \frac{dp_x(g(y))}{dy}\,|g'(y)| + p_x(g(y))\frac{d\,|g'(y)|}{dy} \\ &= \frac{dp_x(g(y))}{dg(y)}\frac{dg(y)}{dy}\,|g'(y)| + p_x(g(y))\frac{d\,|g'(y)|}{dy}. \end{aligned} \tag{1}$$

Hence, if $x_* = g(y_*)$, the

$$\frac{dp_x(g(y_*))}{dg(y_*)} = \frac{dp_x(x_*)}{dx_*} = 0.$$

However, there is no guarantee that the second term of the RHS of Eq. 1 is zero. For example, if $p_x(x) = 2x$ for $0 \le x \le 1$ and $x = \sin(y)$, where $0 \le y \le \pi/2$. Then according to the transformation formula, we have that

$$p_y(y) = p_x(g(y))g'(y) = 2\sin(y)\cos(y) = \sin(2y) \text{ for } 0 \le y \le \frac{\pi}{2}.$$

Clearly, $p_y(y)$ reaches its peak at $y = \pi/4$ but $\sin(\pi/4) \ne x_* = 1$. Thus, we have found a counterexample.

On the other hand, if $g(y)$ is an affine map, then $g'(y)$ is a constant map and as a result

$$\frac{d\,|g'(y)|}{dy} = 0$$

Problem 1.5 [characterization of variance]

It suffices to show that $\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ since any a measurable function of a random variable is again a random variable and in this case f although is not mentioned, it is safe to assume in this context that f is measurable. So note

$$\begin{aligned}\mathrm{Var}[X] &= \mathbb{E}[X - \mathbb{E}[X]]^2 \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + (\mathbb{E}[X])^2 \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2\end{aligned}$$

as desired.

Problem 1.6 [covariance of two independent r.v. is zero]

Since $X \perp Y$, then it follows that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. Then we have

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY - X\mathbb{E}[Y] - \mathbb{E}[X]Y + \mathbb{E}[X]\mathbb{E}[Y]]\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\
&= 0.
\end{aligned}
$$

Problem 1.7 [gaussian integral via polar coordinate]

First, we write

$$
\begin{aligned}
I^2 &= \left(\int_{\mathbb{R}} \exp\left\{ -\frac{1}{2\sigma^2}x^2 \right\} dx\right)\left(\int_{\mathbb{R}} \exp\left\{ -\frac{1}{2\sigma^2}y^2 \right\} dy\right) \\
&= \int\int_{\mathbb{R}\times\mathbb{R}} \exp\left\{ -\frac{1}{2\sigma^2}\left(x^2+y^2\right) \right\} dxdx.
\end{aligned}
$$

Now using polar coordinate - let $x = r\cos\theta$ and $y = r\sin\theta$. Then we get the Jacobian matrix as

$$
\frac{\partial(x,y)}{\partial(r,\theta)} = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix} \implies \left|\frac{\partial(x,y)}{\partial(r,\theta)}\right| = r(\cos\theta^2 + \sin\theta^2) = r.
$$

Hence, as a result

$$
\begin{aligned}
I^2 &= \int_0^{2\pi}\int_0^{\infty} \exp\left\{ -\frac{r^2}{2\sigma^2} \right\} rdrd\theta \\
&=_1 \int_0^{2\pi}\int_0^{\infty} \exp(-u)\sigma^2 dud\theta \\
&= \int_0^{2\pi} \sigma^2 d\theta \int_0^{\infty} \exp(-u)du \\
&= 2\pi\sigma^2\left[-\exp(-u)\right]_0^{\infty} = 2\pi\sigma^2.
\end{aligned}
$$

Problem 1.8 [second moment of gaussian integral via Feymann's trick]

The differentiation under the integral needs a bit more theoretical justification. We won't reproduce the related theorems here. But they could be found in e.g. Theorem 3.2, Theorem 3.3 in Chapter XIII of [Lan97] or in [Con00]. With this in mind, we get

$$
\begin{aligned}
\frac{d}{d\sigma^2}\int_{\mathbb{R}} \exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\} dx &= \int_{\mathbb{R}} \frac{d}{d\sigma^2} \exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\} dx \\
&= \int_{\mathbb{R}} \exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\} (x-\mu)^2 \left(-\frac{1}{2}\right)(\sigma^{-2})^2 dx
\end{aligned}
$$

On the the other hand, we have

$$
\frac{d}{d\sigma^2}(2\pi\sigma^2)^{1/2} = -\frac{1}{2}(2\pi)(\sigma^2)^{-1/2}.
$$

So combined together, we get

$$\int_{\mathbb{R}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}(x-\mu)^2\left(-\frac{1}{2}\right)(\sigma^{-2})^2 dx = \left(-\frac{1}{2}\right)(2\pi)^{1/2}(\sigma^2)^{-1/2}.$$

One step of reduction, we get

$$\begin{aligned}\mathbb{E}[(x-\mathbb{E}[x])^2] &= \mathrm{Var}[x] \\ &= \frac{1}{(2\pi\sigma^2)^{1/2}}\int_{\mathbb{R}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}(x-\mu)^2 dx \\ &= \sigma^2.\end{aligned}$$

And as a result,

$$\mathbb{E}[x^2] = \mathrm{Var}[x] + (\mathbb{E}[x])^2 = \sigma^2 + \mu^2.$$

Problem 1.9 [gaussian density peaks at mean]

It suffices to show the result holds in the multidimensional case since 1-dim is just a special case. Recall that the density of the Gaussian distribution in D dimension is

$$N(x|u,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}.$$

Differentiate w.r.t. x and we get:

$$\nabla_x N(x|u,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}\nabla_x\left(\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right).$$

Now note that $\varphi(x) = (x-\mu)^T\Sigma^{-1}(x-\mu)$ for $x \in \mathbb{R}^d$, then note for any $h \in \mathbb{R}^D$

$$\begin{aligned}\varphi(x+h) &= (x-u+h)^T\Sigma^{-1}(x-\mu+h) \\ &= (x-\mu)^T\Sigma^{-1}(x-\mu+h) + h^T\Sigma^{-1}(x-\mu+h) \\ &= (x-\mu)^T\Sigma^{-1}(x-\mu) + (x-\mu)^T\Sigma^{-1}h + h^T\Sigma^{-1}(x-\mu) + h^T\Sigma^{-1}h \\ &= (x-\mu)^T\Sigma^{-1}(x-\mu) + \left\langle 2\Sigma^{-1}(x-\mu), h\right\rangle + h^T\Sigma^{-1}h\end{aligned}$$

Note that and

$$h^T\Sigma^{-1}h = \left\langle h\Sigma^{-1/2}, h\Sigma^{-1/2}\right\rangle \leq \left\|h\Sigma^{-1/2}\right\|^2 \leq C\|h\|^2\|\Sigma\|_\infty^2 = o(\|h\|),$$

and that $\left\langle 2\Sigma^{-1}(x-\mu), h\right\rangle \in \mathrm{Hom}(\mathbb{R}^d, \mathbb{R})$. It follows that

$$\nabla_x\varphi(x) = 2\Sigma^{-1}(x-\mu),$$

whence

$$\nabla_x\varphi(x) = 0 \iff 2\Sigma^{-1}(x-\mu) = 0 \iff x = \mu.$$

Problem 1.10 [linearity of expectation and variance]

1. Note

$$\begin{aligned}
\mathbb{E}\left[x+y\right] &= \int_{\mathrm{supp}(x)}\int_{\mathrm{supp}(y)}(x+y)f_{(x,y)}(x,y)dxdy\\
&= \int_{\mathrm{supp}(x)}\int_{\mathrm{supp}(y)}(x+y)f_x(x)f_y(y)dxdy\\
&= \int_{\mathrm{supp}(x)}\int_{\mathrm{supp}(y)}xf_x(x)f_y(y)dxdy+\int_{\mathrm{supp}(x)}\int_{\mathrm{supp}(y)}yf_x(x)f_y(y)dxdy\\
&= \int_{\mathrm{supp}(x)}xf_x(x)dx\int_{\mathrm{supp}(y)}f_y(y)+\int_{\mathrm{supp}(x)}f_x(x)dx\int_{\mathrm{supp}(y)}yf_y(y)\\
&= \mathbb{E}[x]+\mathbb{E}[y].
\end{aligned}$$

2. Note

$$\begin{aligned}
\mathrm{Var}[x+y] &= \mathbb{E}[x+y]^2-(\mathbb{E}[x+y])^2\\
&= \mathbb{E}[x^2]+\mathbb{E}[y^2]+\underbrace{2\mathbb{E}[xy]}_{\mathbb{E}[x]\mathbb{E}[y]}-(\mathbb{E}[x])^2-(\mathbb{E}[y])^2-2\mathbb{E}[x]\mathbb{E}[y]\\
&= \mathbb{E}[x^2]-(\mathbb{E}[x])^2+\mathbb{E}[y^2]-(\mathbb{E}[y])^2\\
&= \mathrm{Var}[x]+\mathrm{Var}[y].
\end{aligned}$$

Problem 1.11 [MLE of gaussian]

Recall that the log-likelihood function for Gaussian distribution is

$$\ln p(x|\mu,\sigma^2) = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_n-\mu)^2-\frac{N}{2}\ln\sigma^2-\frac{N}{2}\ln(2\pi).$$

Now we differentiate it w.r.t. μ and setting it to zero:

$$\frac{\partial\ln p(x|\mu,\sigma^2)}{\partial\mu} = -\frac{1}{2\sigma^2}\cdot 2\cdot\sum_{i=1}^{N}(x_n-\mu)=0 \iff \sum_{i=1}^{N}(x_n-\mu)=0 \iff \mu_{ML}=\frac{1}{n}\sum_{i=1}^{N}x_n.$$

Now we differentiate it w.r.t. σ^2 and setting it to zero:

$$\frac{\partial\ln(p|\mu,\sigma^2)}{\partial\sigma^2} = \underbrace{\sum_{n=1}^{N}(x_n-\mu)^2\left(-\frac{1}{2}\right)(-1)(\sigma^2)^{-2}-\frac{N}{2\sigma^2}=0}_{(\star)}.$$

To rearrange, we get

$$(\star) \iff \sum_{n=1}^{N}(x_n-\mu)^2\sigma^{-4}=\frac{N}{\sigma^2}$$

$$\iff \sum_{n=1}^{N}(x_n-\mu)^2 = \sigma^2 N$$

$$\iff \sigma_{ML}^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2.$$

Plug in $\mu = \mu_{ML}$ we get $\sigma_{ML}^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2$ as desired.

Problem 1.12 [inconsistency gaussian MLE]

Problem 1.14 [independent terms of 2-nd order term in polynomial]

We rewrite the sum in matrix form: $\sum_{i=1}^{D}\sum_{j=1}^{D} w_{ij}x_ix_j = x^TWx$, where $[W]_{ij} = w_{ij}$. Define

$$W_S = \frac{1}{2}(W+W^T) \text{ and } W_A = \frac{1}{2}(W-W^T).$$

Clearly, W_S is symmetric and $W_A^T = \frac{1}{2}(W^T - W) = -W_A$ is anti-symmetric and $W_S + W_A = W$. Therefore,

$$x^TWx = x^T(W_S+W_A)x = x^TW_Sx + x^TW_Ax.$$

Notice that

$$x^TW_Ax = \frac{1}{2}(x^TW_Sx - x^TW^Tx) = \frac{1}{2}(x^TW_Ss - x^TWx) = 0,$$

where the last inequality follows from the fact that x^TW^Tx is a scalar and is equal to x^TWx. Since we have shown the sum, $\sum_{i,j} w_{ij}x_ix_j$, only depends on a symmetric matrix, W_S, whose independent items is of the cardinality of $\sum_{i=1}^{D} i = D(D+1)/2$ if we assume its of dimension $D \times D$, we have established our claim.

Problem 1.15 [independent terms of M-th order term in polynomial]

1. Since by writing the M-th order in the form of

$$\sum_{i_1=1}^{D}\sum_{i_2=1}^{D}\cdots\sum_{i_M=1}^{D} w_{i_1,i_2,\cdots,i_M}x_{i_1}x_{i_2}\cdots x_{i_M}$$

introduces duplicate terms, e.g. if $w_{1,3,2}x_1x_3x_2$ and $w_{2,3,1}x_2x_3x_1$ are the same and can be combined into $(w_{1,3,2}+w_{2,3,1})x_1x_2x_3$, we can introduce an ordering that prevents such duplication from happening. Rewrite the sum in the newly introduced ordering yields

$$\sum_{i_1=1}^{D}\sum_{i_2=1}^{i_1}\cdots\sum_{i_M=1}^{i_{M-1}} w_{i_1,i_2,\cdots,i_M}x_{i_1}x_{i_2}\cdots x_{i_M}.$$

Thus, we have

$$\begin{aligned}
n(D,M) &= \sum_{i_1=1}^{D}\sum_{i_2=1}^{i_1}\cdots\sum_{i_M=1}^{i_{M-1}} w_{i_1,i_2,\cdots,i_M} x_{i_1}x_{i_2}\cdots x_{i_M} \\
&= \sum_{i_1=1}^{D}\left(\sum_{i_2=1}^{i_1}\cdots\sum_{i_M=1}^{i_{M-1}} w_{i_1,i_2,\cdots,i_M} x_{i_1}x_{i_2}\cdots x_{i_M}\right) \\
&= \sum_{i_1=1}^{D} n(i_1, M-1).
\end{aligned}$$

2. To show the equality holds using induction, we note for the base case of $D = 1$,

$$\text{LHS} = \frac{(1+M-2)!}{0!(M-1)!} = \frac{(M-1)!}{(M-1)!} = 1.$$

And

$$\text{RHS} = \frac{(1+M-1)!}{(D-1)!M!} = \frac{M!}{M!} = 1.$$

Now suppose $D = k$ and the equality holds. Then

$$\begin{aligned}
\sum_{i=1}^{k+1}\frac{(i+M-2)!}{(i-1)!(M-1)!} &= \sum_{i=1}^{k}\frac{(i+M-2)!}{(i-1)!(M-1)!} + \frac{(k+1+M-2)!}{k!(M-1)!} \\
&= \frac{(k+M-1)!}{(k-1)!M!} + \frac{(k+M-1)!}{k!(M-1)!} \qquad (1) \\
&= \frac{(k+M-1)!(k+M)}{k!(M-1)!} \\
&= \frac{(k+M)!}{k!M!} \\
&= \frac{((k+1)+M-1)!}{(k+1-1)!M!},
\end{aligned}$$

where Eq. (1) follows from induction hypothesis.

3. We establish the identity by inducting on M. By Problem 1.14, it follows that

$$n(D,2) = \frac{1}{2}D(D+1) = \frac{(D+2-1)!}{(D-1)!2!} = \frac{(D+1)!}{(D-1)!2!},$$

which proves the base case. Now suppose the statement holds for $M = k$. Then for $M = k+1$, we have

$$n(D,k+1) = \sum_{i=1}^{D} n(i,k) = \sum_{i=1}^{D}\frac{(i+M-2)!}{(i-1)!(M-1)!} = \frac{(D+M-1)!}{(D-1)!M!}$$

using part-2.

Problem 1.16 [independent terms of high order polynomial]

1. The first equality just follows from that summing up all the independent terms:

$$N(D,M) = \sum_{i=0}^{M} n(D,i).$$

2. We prove this inequality by inducting on M. Now for the base case, $M = 0$, we note that

$$\text{LHS} = n(D,0) = \frac{(D+0-1)!}{(D-1)!0!} = 1 = \frac{(D+0)!}{D!0!} = \text{RHS}.$$

Now assume that the claim holds for $M = k$. Then for $M = k+1$, we have

$$\begin{aligned} N(D,k+1) &= \sum_{i=0}^{k} n(D,i) + n(D,k+1) \\ &= \frac{(D+k)!}{D!k!} + \frac{(D+k+1-1)!}{(D-1)!(k+1)!} \\ &= \frac{(D+k)!(D+k+1)}{D!(k+1)!} \\ &= \frac{(D+k+1)!}{D!(k+1)!}, \end{aligned}$$

proving the inducting step.

3. Now we show that $N(D,M)$ grows in polynomial fashion like D^M. Assume $D \ll M$. First, we write

$$\begin{aligned} N(D,M) &= \frac{(D+M)!}{D!M!} \\ &\simeq \frac{(D+M)^{D+M}e^{-(D+M)}}{D!M^M e^{-M}} && \text{(by Stirling's approximation)} \\ &= \frac{1}{D!M^M}\left(1+\frac{D}{M}\right)^{D+M} M^{D+M}\frac{e^{-(D+M)}}{e^{-M}} \\ &= \frac{e^{-D}}{D!}\left(1+\frac{D}{M}\right)^{D+M} M^D. && (1) \end{aligned}$$

Now we take a more delicate look at the term $(1+\frac{D}{M})^{D+M}$. Note that

$$\begin{aligned} \left(1+\frac{D}{M}\right)^{D+M} &= \left(1+\frac{D}{M}\right)^{M}\left(1+\frac{D}{M}\right)^{D} \\ &= \left(\left(1+\frac{1}{M/D}\right)^{M/D}\right)^{D}\left(1+\frac{D}{M}\right)^{D} \\ &\le e^D 2^D, \end{aligned}$$

where the inequality comes from the fact that $(1+1/x)^x$ is an increasing function and $D <$

$M \Rightarrow D/M \leq 1$. Substitution back into Eq (1), we get

$$N(D,M) \leq \frac{e^{-D}}{D!} e^D 2^D M^D = \frac{2^D}{D!} M^D.$$

The case for $M \ll D$ follows by symmetry.

Problem 1.17 [gamma density warmup]

1. Note

$$\begin{aligned}\Gamma(x+1) &= \int_0^\infty u^x e^{-u} du \\ &= \left[-u^x e^{-u}\right]_{u=0}^\infty + \int_0^\infty x u^{x-1} e^{-u} du \\ &= x\Gamma(x).\end{aligned}$$

2. We note that

$$\Gamma(1) = \int_0^\infty e^{-u} du = \left[e^{-u}\right]_0^\infty = 1.$$

And as a result, by recursion

$$\Gamma(x+1) = x\Gamma(x) = \cdots = x! \text{ for } x \in \mathbb{N}.$$

Problem 1.18 [volume of unit sphere in n-space]

To state the problem statement in a clearer manner, we solve this problem in several steps. In this problem, we let $d\mu$ denote the Lebesgue measure.

1. First we derive Eq (1.142) in the book. We first rewrite the LHS in the following way. Let $x \in \mathbb{R}^D$ be arbitrary, then

$$\begin{aligned}\int_{\mathbb{R}^d} e^{-\|x\|^2} dx &= \int_{\mathbb{R}} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} e^{-(x_1^2 + x_2^2 + \cdots + x_n^2)} dx_1 dx_2 \cdots dx_n \\ &= \prod_{i=1}^D \int_{\mathbb{R}} e^{-x_i^2} dx_i.\end{aligned}$$

Next, we evaluate this integral. In order to make the computation easier, we choose to let the integrand be $e^{-\pi|x|^2}$instead (it doesn't effect the final result, and one could always get the original integral by scaling). Note that using the same argument as above, we have

$$\int_{\mathbb{R}^D} e^{-\pi\|x\|^2} dx = \left(\int_{\mathbb{R}} e^{-\pi x^2} dx\right)^D.$$

Next, we have

$$\left(\int_{\mathbb{R}} e^{-\pi x^2} dx\right)^2 = \left(\int_{\mathbb{R}} e^{-\pi x_1^2} dx_1\right)\left(\int_{\mathbb{R}} e^{-\pi x_2^2} dx_2\right)$$

$$
\begin{aligned}
&= \int_{\mathbb{R}\times\mathbb{R}} e^{-\pi(x_1^2+x_2^2)} d(x_1 \times x_2) && \text{(by Fubini's theorem)} \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-\pi(x_1^2+x_2^2)} dx_1 dx_2 && \text{(by Fubini's theorem)} \\
&= \int_{[0,2\pi]} \int_{\mathbb{R}} e^{-\pi r^2} r dr d\theta && \text{(switch to polar coordinates)} \\
&= \int_{[0,2\pi]} d\theta \int_{\mathbb{R}} e^{-\pi r^2} r dr \\
&= 2\pi \left[-\frac{1}{2\pi} e^{-\pi r^2} \right]_0^\infty \\
&= 1.
\end{aligned}
$$

Since $\int_{\mathbb{R}} e^{\pi x^2} dx > 0$, it follows that $\int_{\mathbb{R}^D} e^{-\pi\|x\|^2} dx = 1$.

2. Consider the function $f : \mathbb{R}^D \to \mathbb{R}; x \mapsto e^{-\pi\|x\|^2}$. We just showed in part-1 that $f \in L^1(\mathbb{R}^D)$. Therefore, using generalized spherical coordinate (e.g. Theorem 6.3.4 in [Ste05]), we have that

$$
\begin{aligned}
1 = \int_{\mathbb{R}^D} f(x)dx &= \int_{S^{D-1}} \left(\int_{\mathbb{R}^+} f(r\gamma) r^{D-1} dr \right) d\sigma(\gamma) \\
&= \int_{S^{D-1}} \left(\int_{\mathbb{R}^+} e^{-\pi\|r\gamma\|^2} r^{D-1} dr \right) d\sigma(\gamma) \\
&= \int_{S^{D-1}} \left(\int_{\mathbb{R}^+} e^{-\pi r^2} r^{D-1} dr \right) d\sigma(\gamma) \\
&= \int_{S^{D-1}} d\sigma(r) \int_{\mathbb{R}^+} e^{-\pi r^2} r^{D-1} dr \\
&= \sigma(S^{D-1}) \int_{\mathbb{R}^+} e^{-\pi r^2} r^{D-1} dr.
\end{aligned}
$$

Now we evaluate the integral on the RHS:

$$
\begin{aligned}
\int_{\mathbb{R}^+} e^{-\pi r^r} r^{D-1} dr &= \int_0^\infty e^{-u} \left(\frac{u}{\pi}\right)^{\frac{D-1}{2}} \frac{1}{2\pi(u/\pi)^{1/2}} du \\
&= \frac{1}{2\pi} \int_0^\infty e^{-u} \left(\frac{u}{\pi}\right)^{\frac{D}{2}-1} du \\
&= \frac{1}{2\pi} \pi^{1-\frac{D}{2}} \int_0^\infty e^{-u} u^{\frac{D}{2}-1} du \\
&= \frac{1}{2} \pi^{-\frac{D}{2}} \Gamma\left(\frac{D}{2}\right).
\end{aligned}
$$

Therefore, substituting back we get

$$
\sigma(S^{D-1}) = \frac{1}{\int_{\mathbb{R}^+} e^{-\pi r^2} r^{D-1} dr} = \frac{2\pi^{D/2}}{\Gamma(D/2)}.
$$

This $\sigma(S^{D-1})$ is the S_D in the problem.

3. Now we calculate the volume of the ball. Let B_1 denote the unit ball in $\mathbb{R}^D$. Note that again

by generalized spherical coordinate,

$$
\begin{aligned}
V_D &= \int_{\mathbb{R}^D} \mathbb{1}_{B_1}(x) d\mu \\
&= \int_{S^{D-1}} \int_{\mathbb{R}^+} \mathbb{1}_{B_1}(r\gamma) r^{D-1} d\sigma(\gamma) \\
&= \int_{S^{D-1}} \left(\int_{[0,1]} r^{D-1} dr \right) d\sigma(\gamma) \\
&= \left(\int_{S^{D-1}} d\sigma(\gamma) \right) \left(\int_{[0,1]} r^{D-1} dr \right) \\
&= \sigma(S^{D-1}) \left[\frac{1}{D} r^D \right]_0^1 \\
&= \frac{\pi^{D/2}}{\Gamma(D/2)(D/2)} \\
&= \frac{\pi^{D/2}}{\Gamma(D/2+1)}.
\end{aligned}
$$

as desired.

4. When $D = 2$, we get
$$S_D = \frac{2\pi^{2/2}}{\Gamma(1)} = 2\pi \text{ and } V_D = \frac{S_D}{D} = \pi.$$
When $D = 2$, we get
$$S_D = \frac{2\pi^{3/2}}{\Gamma(3/2)} = \frac{2\pi^{3/2}}{\pi^{1/2}/2} = 4\pi \text{ and } V_D = \frac{4}{3}\pi.$$

Remark 1.1. This problem could have been solved heuristically. But it loses rigor. What was showed was a rigorous mathematical way to treat this problem.

Problem 1.19 [high dimensional cubes concentrate on corners]

1. Using the result of the previous problem, and the fact that $m_d(rB) = r^d m(B)$, where m_d is the Lebesgue measure in d-dimensional Euclidean space (e.g. Exercise 1.6 in [Ste05]), we have that

$$
\begin{aligned}
\frac{V_{\text{sphere}}}{V_{\text{cube}}} &= \frac{\pi^{D/2} a^D}{\Gamma(D/2+1) 2^D a^D} = \frac{\pi^{D/2}}{\Gamma(D/2+1) 2^D} \\
&\simeq \frac{\pi^{D/2}}{(2\pi)^{1/2} e^{-D/2} (D/2)^{D/2+1/2} 2^D} && \text{(by Stirling formula)} \\
&= C \frac{\pi^{D/2} e^{D/2}}{(D/2)^{D/2}} \frac{1}{D^{1/2}} 2^{-D} && \text{(C is some constant)} \\
&= C \left(\frac{2\pi e}{D} \right)^{D/2} \frac{1}{D^{1/2} 2^D} \xrightarrow{D \to \infty} 0.
\end{aligned}
$$

2. On the other hand, we have

$$\begin{aligned}\text{dist(center to corner)} &= \sqrt{Da^2} = a\sqrt{D} \\ \text{dist(center to top)} &= a.\end{aligned}$$

   And thus the ratio is $\sqrt{D}$.

Problem 1.20 [high dimensional gaussian concentrate on strip]

Bibliography

C

[Con00] Keith Conrad. Differentiating under the integral sign. 2000. 5

L

[Lan97] Serge Lang. *Undergraduate Analysis*. Springer-Verlag New York, 2 edition, 1997. 5

S

[Ste05] Elias Stein. *Real analysis : measure theory, integration, and Hilbert spaces*. Princeton University Press, Princeton, N.J. Oxford, 2005. 12, 13