演習問題 1.20

この演習問題では、高次元ガウス分布の振る舞いを扱う。D 次元ガウス分布

$$p(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{D/2}} \exp\left(-\frac{\|\mathbf{x}\|^2}{2\sigma^2}\right) \quad \cdots (1.47)$$

を考えよう。極座標系で、角度方向については積分し、半径に関する密度を求めたい。このために、半径 r にある厚さ ϵ の薄皮に関して $\mathbf{x}$ の確率密度の積分をとると、$\epsilon \ll 1$ のとき、$p(r)\epsilon$ となることを示せ。ただし、

$$p(r) = \frac{S_D r^{D-1}}{(2\pi\sigma^2)^{D/2}} \exp\left(-\frac{\|r\|^2}{2\sigma^2}\right) \quad \cdots (1.48)$$

であり、S_D は D 次元単位球の表面積である。関数 $p(r)$ が 1 つの停留点をもち、D が大きいとき、$\hat{r} \cong \sqrt{D}\,\sigma$ になることを示せ。$\epsilon \ll \hat{r}$ について $p(\hat{r}+\epsilon)$ を考え、D が大きいとき、

$$p(\hat{r}+\epsilon) = p(\hat{r}) \exp\left(-\frac{\epsilon^2}{\sigma^2}\right) \quad \cdots (1.49)$$

となることを示せ。これから、$\hat{r}$ が確率密度の最大を与える半径となり、$p(r)$ が $\hat{r}$ での最大値から σ の長さのスケールで指数的に減衰していることがわかる。我々はすでに、D が大きいとき、$\sigma \ll \hat{r}$ であることを見てきたので、ほとんどの確率質量が大きな半径の薄皮に集中していることがわかる。最後に、$\mathbf{x}$ の確率密度 $p(\mathbf{x})$ は、半径 $\hat{r}$ にある地点よりも、原点での方が $\exp(D/2)$ 倍大きいことを示せ。このことから、ほとんどの高次元ガウス分布の確率質量は、$\mathbf{x}$ の確率密度の高いところとは異なる半径のところにあることがわかる。後の章でモデルパラメータのベイズ推論を考える際に、高次元空間の分布のこの性質を用いて、重要な結論を導くことになる。

[解]

まず、半径 r にある厚さ ϵ の薄皮に関して $\mathbf{x}$ の確率密度の積分をとると、$\epsilon \ll 1$ のとき、$p(r)\epsilon$ となることを示す。式 (1.48) より、$p(r)$ は

$$p(r) = \frac{S_D r^{D-1}}{(2\pi\sigma^2)^{D/2}} \exp\left(-\frac{\|r\|^2}{2\sigma^2}\right)$$

で表されるので、r から $r+\epsilon$ の確率密度の積分を求めると、その累積分布関数 $P(r)$ は以下のように表される。

$$P(r) = \int_r^{r+\epsilon} p(r) = \frac{S_D}{(2\pi\sigma^2)^{D/2}} \int_r^{r+\epsilon} e^{\left(-\frac{\|r\|^2}{2\sigma^2}\right)} r^{D-1}\,dr$$

ここで、$f(r) = e^{\left(-\frac{\|r\|^2}{2\sigma^2}\right)} r^{D-1}$ とし、この不定積分の原始関数 $F(r)$ をとすると、上記の式は以下のようになる。

$$= \frac{S_D}{(2\pi\sigma^2)^{D/2}} [\ F(r)\ ]_r^{r+\epsilon}$$

$$= \frac{S_D}{(2\pi\sigma^2)^{D/2}} (\ F(r+\epsilon) - F(r)\ )$$

ここで、ϵ は微小区間なので、

$$= \frac{S_D}{(2\pi\sigma^2)^{D/2}} \cdot \frac{F(r+\epsilon) - F(r)}{\epsilon} \cdot \epsilon$$

$$= \frac{S_D}{(2\pi\sigma^2)^{D/2}} \cdot \lim_{\epsilon \to 0} F(r) \cdot \epsilon$$

$$= \frac{S_D}{(2\pi\sigma^2)^{D/2}} \cdot F'(r) \cdot \epsilon$$

$$= \frac{S_D}{(2\pi\sigma^2)^{D/2}} \cdot f(r) \cdot \epsilon$$

$$= \frac{S_D}{(2\pi\sigma^2)^{D/2}} \cdot e^{\left(-\frac{\|r\|^2}{2\sigma^2}\right)} r^{D-1} \cdot \epsilon$$

となり、式 (1.48) であるので、上記の式は、

$$= p(r)\epsilon$$

と変形できる。よって、半径 r にある厚さ ϵ の薄皮に関して $\mathbf{x}$ の確率密度の積分をとると、$\epsilon \ll 1$ のとき、$p(r)\epsilon$ となることを示せた。

[**微分係数**]

定義

関数 $f(x)$ の $x = a$ における微分係数

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} = \lim_{h \to 0} \frac{f(x) - f(a)}{x - a}$$

微分可能と連続

関数 $f(x)$ の $x = a$ において微分可能ならば、関数 $f(x)$ は $x = a$ で連続である。ただし、逆は成り立たない。

[**導関数**]

定義

関数 $f(x)$ の導関数

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

次に、関数 $p(r)$ が 1 つの停留点をもち、D が大きいとき、$\hat{r} \cong \sqrt{D}\,\sigma$ になることを示す。
停留点は微分した値が 0 になるところであるので、式 (1.48) を

$$\frac{\partial\, p(r)}{\partial\, r} = \left\{ \frac{S_D\, r^{D-1}}{(2\pi\sigma^2)^{D/2}}\, e^{\left(-\frac{\| r \|^2}{2\sigma^2}\right)} \right\}'$$

※注 … ここで r はベクトルではないので、ノルムは外れ、$\| r \|^2 = r^2$ と書き直せる。

$$= \left\{ \frac{S_D}{(2\pi\sigma^2)^{D/2}}\, r^{D-1} \right\}' e^{\left(-\frac{r^2}{2\sigma^2}\right)} + \frac{S_D}{(2\pi\sigma^2)^{D/2}}\, r^{D-1} \left\{ e^{\left(-\frac{r^2}{2\sigma^2}\right)} \right\}'$$

$$= (D-1)\frac{S_D}{(2\pi\sigma^2)^{D/2}}\, r^{D-2}\, e^{\left(-\frac{r^2}{2\sigma^2}\right)} + \frac{S_D}{(2\pi\sigma^2)^{D/2}}\, r^{D-1} \left(-\frac{r^2}{2\sigma^2}\right)' e^{\left(-\frac{r^2}{2\sigma^2}\right)}$$

$$= (D-1)\frac{S_D}{(2\pi\sigma^2)^{D/2}}\, r^{D-2}\, e^{\left(-\frac{r^2}{2\sigma^2}\right)} - \frac{r}{\sigma^2} \cdot \frac{S_D}{(2\pi\sigma^2)^{D/2}}\, r^{D-1}\, e^{\left(-\frac{r^2}{2\sigma^2}\right)}$$

$$= (D-1)\frac{S_D}{(2\pi\sigma^2)^{D/2}}\, r^{D-2}\, e^{\left(-\frac{r^2}{2\sigma^2}\right)} - \frac{r^2}{\sigma^2} \cdot \frac{S_D}{(2\pi\sigma^2)^{D/2}}\, r^{D-2}\, e^{\left(-\frac{r^2}{2\sigma^2}\right)}$$

$$= \left\{ (D-1) - \frac{r^2}{\sigma^2} \right\} \frac{S_D}{(2\pi\sigma^2)^{D/2}}\, r^{D-2}\, e^{\left(-\frac{r^2}{2\sigma^2}\right)} = 0$$

と変形すると、これを 0 にする r の値は、

$$r = 0,\quad \sqrt{D-1}\,\sigma$$

となる。今、半径 r が 0 であることはあり得ないので、r の値は、

$$r = \sqrt{D-1}\,\sigma$$

のただ 1 つの停留点となり、D が大きいとき、

$$\hat{r} \cong \sqrt{D}\,\sigma$$

と近似できることがわかる。よって、関数 $p(r)$ が 1 つの停留点をもち、D が大きいとき、$\hat{r} \cong \sqrt{D}\,\sigma$ となることが示せた。

[積の導関数]

$$(f(x)\,g(x))' = f'(x)\,g(x) + f(x)\,g'(x)$$

さらに、$\epsilon \ll \hat{r}$ について $p(\hat{r}+\epsilon)$ を考え、D が大きいとき、式 (1.49) となることを示す。式 (1.49) の左辺は、$\hat{r} = \hat{r}+\epsilon$ のとき、式 (1.48) より、

$$p(\hat{r}+\epsilon) = \frac{S_D\,(\hat{r}+\epsilon)^{D-1}}{(2\pi\sigma^2)^{D/2}}\, e^{-\frac{(\hat{r}+\epsilon)^2}{2\sigma^2}}$$

となる。ここで、$p(\hat{r}) = \frac{S_D\,\hat{r}^{D-1}}{(2\pi\sigma^2)^{D/2}} \exp\left(-\frac{\hat{r}^2}{2\sigma^2}\right)$ の形を作りたいので、上記の式は、

$$= \frac{S_D\,\hat{r}^{D-1}\left(1+\frac{\epsilon}{\hat{r}}\right)^{D-1}}{(2\pi\sigma^2)^{D/2}}\, e^{\left(-\frac{\hat{r}^2}{2\sigma^2}-\frac{2\hat{r}\epsilon}{2\sigma^2}-\frac{\epsilon^2}{2\sigma^2}\right)}$$

$$= \frac{S_D\,\hat{r}^{\,D-1}}{(\,2\pi\sigma^2\,)^{D/2}}\,e^{-\frac{\hat{r}^2}{2\sigma^2}}\cdot\left(1+\frac{\epsilon}{\hat{r}}\right)^{D-1}\,e^{\left(-\frac{\hat{r}\epsilon}{\sigma^2}-\frac{\epsilon^2}{2\sigma^2}\right)}$$

$$= p(\,\hat{r}\,)\left(1+\frac{\epsilon}{\hat{r}}\right)^{D-1}\,e^{\left(-\frac{\hat{r}\epsilon}{\sigma^2}-\frac{\epsilon^2}{2\sigma^2}\right)}$$

… ※

と変形できる。そして、残りの部分である $\left(1+\frac{\epsilon}{\hat{r}}\right)^{D-1}\,e^{\left(-\frac{\hat{r}\epsilon}{\sigma^2}-\frac{\epsilon^2}{2\sigma^2}\right)}$ を $e^{-\frac{\epsilon^2}{\sigma^2}}$ の形に近似していく。ここで、$\left(1+\frac{\epsilon}{\hat{r}}\right)^{D-1}$ を、底が e である指数関数形式に直してやる。すると、

$$e^x = \left(1+\frac{\epsilon}{\hat{r}}\right)^{D-1}$$

$$\ln e^x = \ln\left(1+\frac{\epsilon}{\hat{r}}\right)^{D-1}$$

$$x = \ln\left(1+\frac{\epsilon}{\hat{r}}\right)^{D-1}$$

より、

$$\left(1+\frac{\epsilon}{\hat{r}}\right)^{D-1} = e^{\ln\left(1+\frac{\epsilon}{\hat{r}}\right)^{D-1}}$$

で表されることから、式 ※ は、

$$= p(\,\hat{r}\,)\cdot e^{\ln\left(1+\frac{\epsilon}{\hat{r}}\right)^{D-1}}\cdot e^{\left(-\frac{\hat{r}\epsilon}{\sigma^2}-\frac{\epsilon^2}{2\sigma^2}\right)}$$

$$= p(\,\hat{r}\,)\cdot e^{\left(-\frac{\hat{r}\epsilon}{\sigma^2}-\frac{\epsilon^2}{2\sigma^2}+\ln\left(1+\frac{\epsilon}{\hat{r}}\right)^{D-1}\right)}$$

$$= p(\,\hat{r}\,)\cdot e^{\left(-\frac{\hat{r}\epsilon}{\sigma^2}-\frac{\epsilon^2}{2\sigma^2}+(\,D-1\,)\ln\left(1+\frac{\epsilon}{\hat{r}}\right)\right)}$$

より、$e^{-\frac{\epsilon^2}{\sigma^2}}$ を取り出して、

$$= p(\,\hat{r}\,)\cdot e^{-\frac{\epsilon^2}{\sigma^2}}\cdot e^{\left(\frac{\epsilon^2}{\sigma^2}-\frac{\hat{r}\epsilon}{\sigma^2}-\frac{\epsilon^2}{2\sigma^2}+(\,D-1\,)\ln\left(1+\frac{\epsilon}{\hat{r}}\right)\right)}$$

$$= p(\,\hat{r}\,)\cdot e^{-\frac{\epsilon^2}{\sigma^2}}\cdot e^{\left(\frac{\epsilon^2}{2\sigma^2}-\frac{\hat{r}\epsilon}{\sigma^2}+(\,D-1\,)\ln\left(1+\frac{\epsilon}{\hat{r}}\right)\right)}$$

$$= p(\,\hat{r}\,)\cdot e^{-\frac{\epsilon^2}{\sigma^2}}\cdot e^{\left(\frac{\epsilon^2-2\hat{r}\epsilon}{2\sigma^2}+(\,D-1\,)\ln\left(1+\frac{\epsilon}{\hat{r}}\right)\right)}$$

$$= p(\,\hat{r}\,)\cdot e^{-\frac{\epsilon^2}{\sigma^2}}\cdot e^{\left(\frac{\epsilon^2-2\hat{r}\epsilon}{2\sigma^2}+(\,D-1\,)\ln\left(1+\frac{\epsilon}{\hat{r}}\right)\right)}$$

$$= p(\hat{r}) \cdot e^{-\frac{\epsilon^2}{\sigma^2}} \cdot e^{\left(\frac{(\varepsilon - \hat{r})^2}{2\sigma^2} - \frac{\hat{r}^2}{2\sigma^2} + (D-1)\ln\left(1 + \frac{\epsilon}{\hat{r}}\right) \right)}$$

… ※※

と整理できる。ここで、今 D が十分に大きいので、先程示した $\hat{r} \cong \sqrt{D}\,\sigma$ より、$\hat{r}$ もまた十分に大きくなることが保証されることと、ϵ は微小区間であることから、

$$\lim_{\varepsilon \to 0} (\varepsilon - \hat{r})^2 = (-\hat{r})^2 = \hat{r}^2$$

$$\lim_{\varepsilon \to 0} \ln\left(1 + \frac{\epsilon}{\hat{r}}\right) = \ln 1 = 0$$

となることがわかる。すると、最終的に式 ※※ は、

$$\cong p(\hat{r}) \cdot e^{-\frac{\epsilon^2}{\sigma^2}} \cdot e^{\left(\frac{\hat{r}^2}{2\sigma^2} - \frac{\hat{r}^2}{2\sigma^2} + 0 \right)}$$

$$= p(\hat{r}) \cdot e^{-\frac{\epsilon^2}{\sigma^2}} \cdot e^0$$

$$= p(\hat{r}) \cdot e^{-\frac{\epsilon^2}{\sigma^2}}$$

となる。以上より、D が大きいとき、式 (1.49) となることを示せた。

最後に、$\mathbf{x}$ の確率密度 $p(\mathbf{x})$ は、半径 $\hat{r}$ にある地点よりも、原点での方が $\exp(D/2)$ 倍大きいことを示す。このために、原点での確率密度と $\hat{r}$ での確率密度の比を求める。式 (1.47) より、原点での確率密度 $p(\|\mathbf{x}\| = 0)$ は、

$$p(\|\mathbf{x}\| = 0) = \frac{1}{(2\pi\sigma^2)^{D/2}}\, e^0 = \frac{1}{(2\pi\sigma^2)^{D/2}}$$

となる。一方、$\hat{r}$ での確率密度 $p(\|\mathbf{x}\| = r)$ は、D が大きいとき、$\hat{r} \cong \sqrt{D}\,\sigma$ になることがわかっているので、

$$p(\|\mathbf{x}\| = r) = \frac{1}{(2\pi\sigma^2)^{D/2}}\, e^{-\frac{\hat{r}^2}{2\sigma^2}}$$

$$= \frac{1}{(2\pi\sigma^2)^{D/2}}\, e^{-\frac{D}{2}}$$

となる。この比をとると、

$$\frac{p(\|\mathbf{x}\| = 0)}{p(\|\mathbf{x}\| = r)} = \frac{\frac{1}{(2\pi\sigma^2)^{D/2}}}{\frac{1}{(2\pi\sigma^2)^{D/2}}\, e^{-\frac{D}{2}}} = e^{\frac{D}{2}}$$

となるので、$\mathbf{x}$ の確率密度 $p(\mathbf{x})$ は、半径 $\hat{r}$ にある地点よりも、原点での方が $\exp(D/2)$ 倍大きいことを示せた。