

汎化リスク

名前:松島完忠
学籍番号:t211d070
日付:7/18

[演習 240] 閾値関数の汎化リスク

1. 定数 $\sigma = 0.5$ おき、分布 D_z から 1000 点 mp データ $(x_1, y_1), \dots, (x_{1000}, y_{1000})$ を生成し、頻度分布のプロット結果を図 1 に示す。

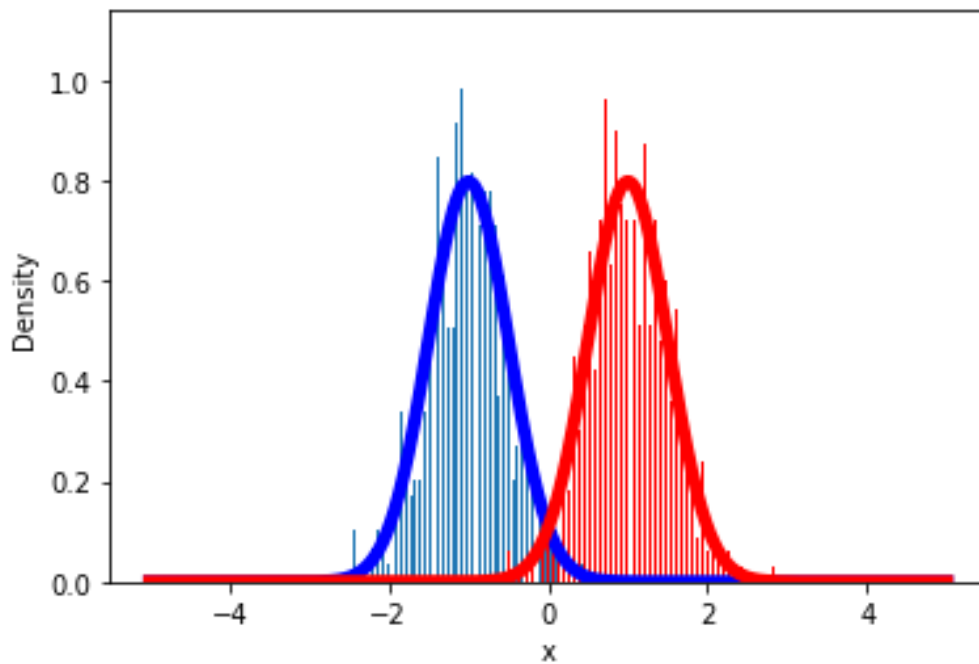


図 1: 分布 D_z の頻度分布

2. $\theta \in \{-5.0, -4.9, \dots, 4.9, 5.0\}$ それぞれに対する閾値関数の汎化リスクを、期待値をプロットした結果を示す。このとき、定数 $\sigma = 0.5$ 、1000000 点のデータを作成し、それ元

に汎化リスクを求めた。

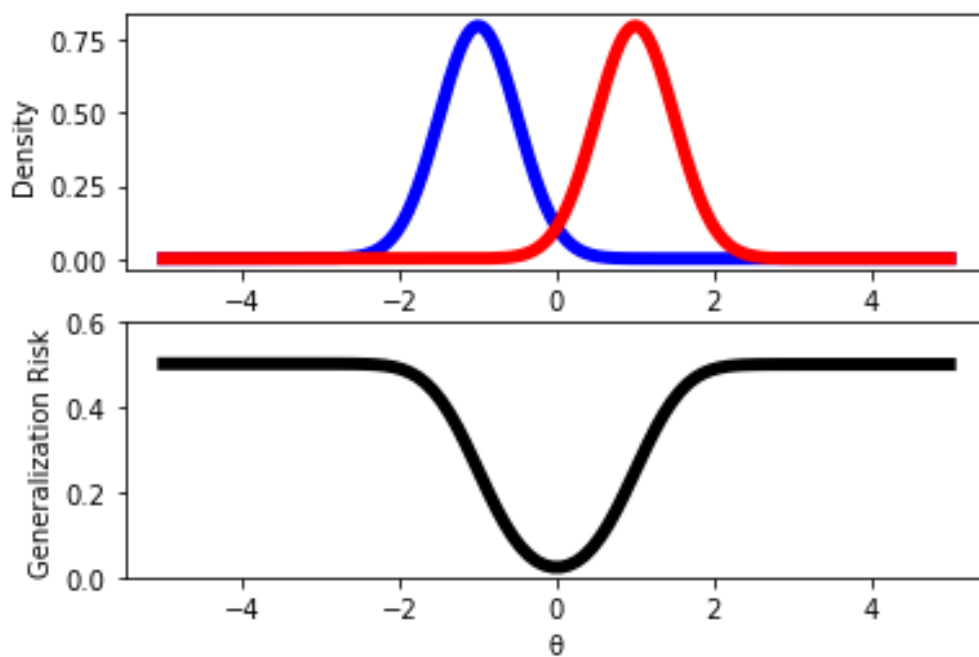


図 2: $\sigma=0.5$ 、経験リスクのプロット

図 2 より陰性データと陽性データの分布関数が重なる部分で経験リスクが 0 に近い値となった。

$3.\sigma = 2, 1, 0.25$ のときの経験リスクを図 3、図 4、図 5 にプロット結果を示す。

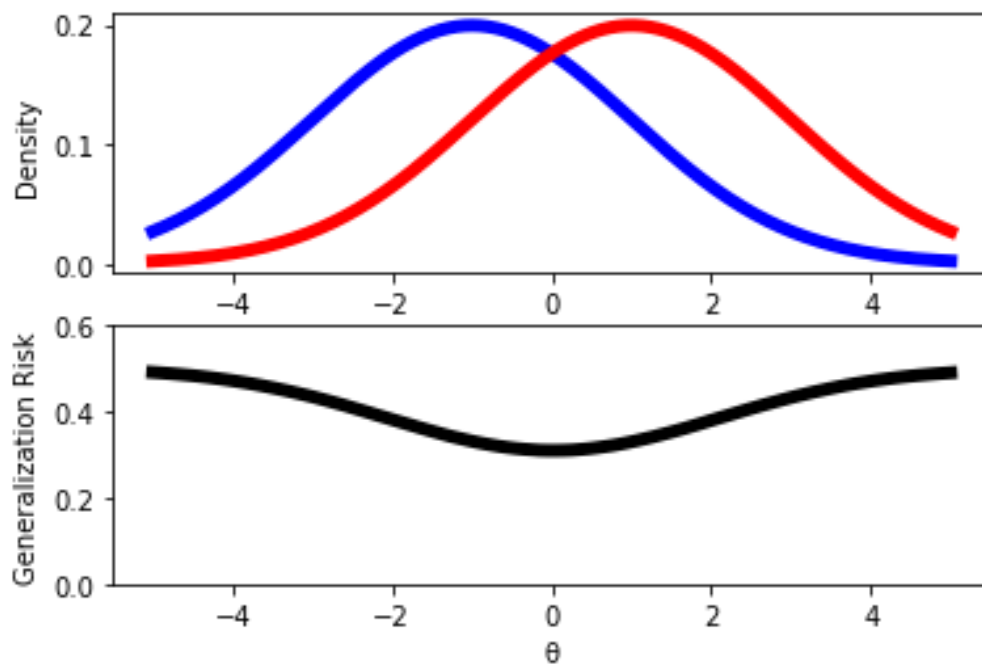


図 3: $\sigma=2$ のときの経験リスク

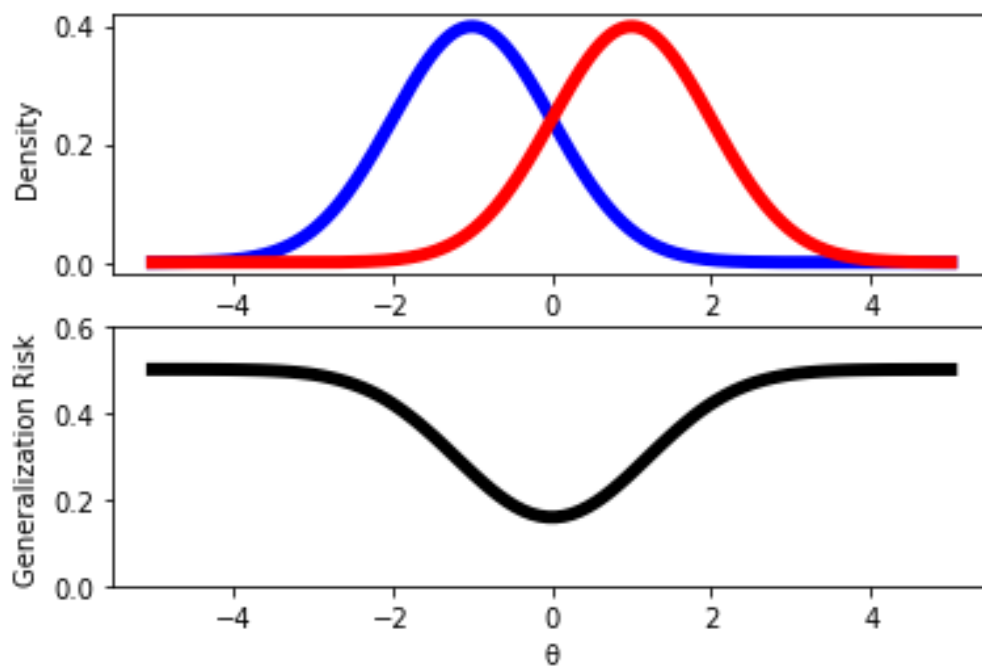


図 4: $\sigma=1$ のときの経験リスク

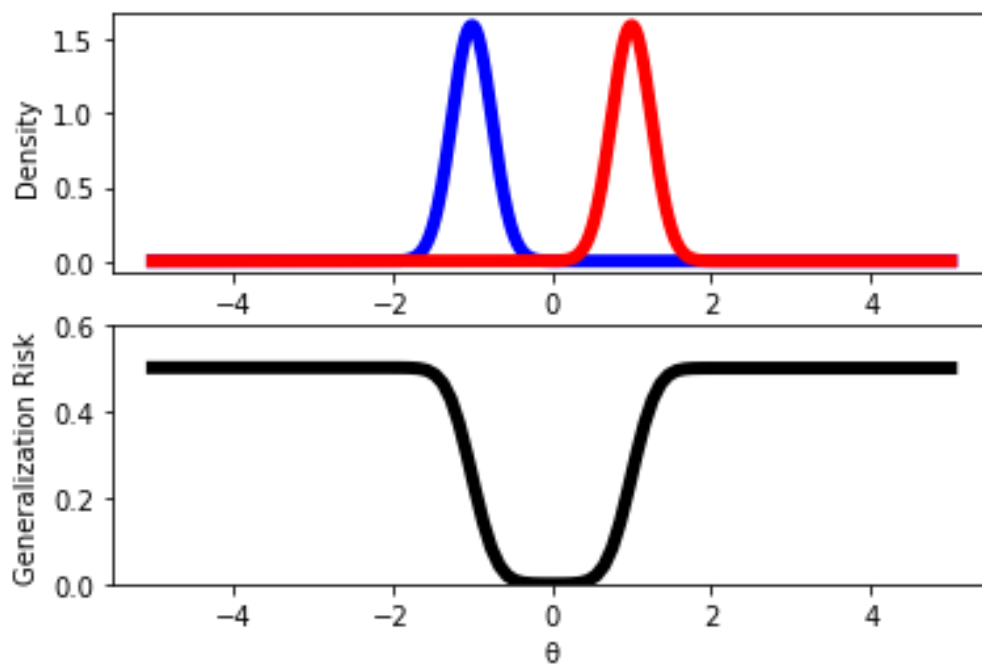


図 5: $\sigma=0.25$ のときの経験リスク

図 3、図 4、図 5 から陽性データと陰性データの分布関数が重なる部分が多いほど
経験リスクの最小値は大きくなる。

作成プログラム

図 6 に作成したプログラムを示す。

1	<code>import numpy as np</code>
2	<code>from scipy import stats</code>
3	<code>import matplotlib.pyplot as plt</code>
4	<code>import scipy.stats</code>
5	<code>import matplotlib.ticker as ticker</code>

6	
7	def Genrisk(a, b, theta, N):
8	sum=0
9	
10	for (x, y) in zip(a, b):
11	f=x-theta
12	if (y*f<=0):
13	sum+=1
14	return float(sum)/float(N)
15	
16	
17	size = 1000000
18	sigma = 2
19	
20	pe=[]
21	ne=[]
22	genrisk=[]
23	x=[]
24	
25	y=stats.uniform.rvs(0, scale=1, size=size)
26	Y=[]
27	
28	
29	for s in y:

30	if s<0.5:
31	t=stats.norm.rvs(loc=-1, scale=sigma)
32	x.append(t)
33	ne.append(t)
34	Y.append(-1.0)
35	else:
36	t=stats.norm.rvs(loc=1, scale=sigma)
37	x.append(t)
38	pe.append(t)
39	Y.append(1.0)
40	
41	
42	xx = np.linspace(-5, 5, 10000)
43	theta = np.linspace(-5, 5, 100)
44	
45	for th in theta:
46	genrisk.append(Genrisk(x, Y, th, size))
47	
48	fig = plt.figure()
49	ax=fig.add_subplot(211, xlabel='x', ylabel='Density')
50	ax.plot(xx, scipy.stats.norm.pdf(xx, -1, sigma), c='blue', linewidth = 5.0)
51	ax.plot(xx, scipy.stats.norm.pdf(xx, 1, sigma), c='red', linewidth = 5.0)
52	
53	

54	
55	<code>ax2=fig. add_subplot (212, xlabel=' θ ', ylabel=' Generalization Risk', ylim=(0, 0.6))</code>
56	<code>ax2. plot (theta, genrisk, c=' black' , linewidth=5.0)</code>

図 6:作成したプログラム