目標函數 $f = ||Y - Y_{pred}||_F^2$

$h = XW_1 + b_1 ; h_{sigmoid} = sigmoid(h) ;$
$Y_{pred} = h_{sigmoid}W_2 + b_2$

表示方式 $h_{sigmoid} = h_s ; sigmoid(h) = S(h)$ $; Y_{pred} = Y_p$

另外來看
$f = ||Y - (S(XW_1 + b_1).W_2 + b_2)||_F^2$

給定實驗資料

In [1]:
```
import torch
import numpy as np
torch.manual_seed(0)

x = torch.randn(100, 1, requires_grad=True)
y = torch.randn(100, 1, requires_grad=True)
w1 = torch.randn(1, 20, requires_grad=True)
w2 = torch.randn(20, 1, requires_grad=True)
b1 = torch.randn(100, 20, requires_grad=True)
b2 = torch.randn(100, 1, requires_grad=True)
print( "x : ", np.shape(x))
print( "y : ", np.shape(y))
print( "w1 : ", np.shape(w1))
print( "w2 : ", np.shape(w2))
print( "b1 : ", np.shape(b1))
print( "b2 : ", np.shape(b2))
```

```
x :  torch.Size([100, 1])
y :  torch.Size([100, 1])
w1 :  torch.Size([1, 20])
w2 :  torch.Size([20, 1])
b1 :  torch.Size([100, 20])
b2 :  torch.Size([100, 1])
```

In [2]:
```
import torch.nn as nn
tm = nn.Sigmoid()
hs = tm(x.mm(w1)+b1)
# print(s)
yp = (hs).mm(w2)+b2
f1 = (y - yp).pow(2).sum()
ft = (y - yp).pow(2)
print(f1)
```

```
tensor(326.1264, grad_fn=<SumBackward0>)
```

In [3]:
```
print( "x : ", x.grad)
print( "y : ", y.grad)
print( "w1 : ", w1.grad)
print( "w2 : ", w2.grad)
print( "b1 : ", b1.grad)
print( "b2 : ", b2.grad)
```

```
x :  None
y :  None
w1 :  None
w2 :  None
b1 :  None
b2 :  None
```

In [4]:
```
f1.backward()
```

直接求導

In [5]:
```
print( "x : ", x.grad)
```

```
x :  tensor([[-1.9933e-01],
        [ 6.7720e-01],
        [-2.0645e+00],
        [-2.8853e-01],
        [ 9.7789e-01],
        [-4.2798e-02],
        [-5.4575e-01],
        [-1.4895e-01],
        [ 4.0874e-02],
        [ 2.2387e-02],
        [ 2.1316e+00],
        [-3.5223e-01],
        [ 7.0717e-01],
        [-4.7275e-01],
        [-9.6598e-01],
        [-4.2666e+00],
        [ 7.7769e-05],
        [ 1.1090e-01],
        [ 9.8039e-01],
        [-2.1306e+00],
        [ 3.4738e-01],
        [-8.7577e-01],
        [ 6.0959e-02],
        [-1.6884e-01],
        [-4.3426e-01],
        [ 1.0328e-01],
        [-3.3121e-01],
        [-6.3263e-01],
        [-2.0504e-01],
        [-1.0108e-01],
        [-1.0438e+00],
        [-4.1394e-01],
        [-1.0305e-02],
        [-1.1881e+00],
        [-2.5335e-01],
        [-3.6720e-01],
        [ 1.2579e-01],
        [-3.4726e+00],
        [ 7.4866e-01],
        [ 4.4494e-01],
```

```
[-1.7558e+00],
[-9.2362e-03],
[-3.5294e-01],
[-8.4580e-01],
[-1.7698e-03],
[ 5.0107e-04],
[-3.0033e-01],
[-6.7953e-02],
[ 6.1309e-02],
[-1.6736e-01],
[ 2.3652e-02],
[-1.4754e+00],
[ 4.4175e-01],
[-6.0080e-01],
[ 8.8950e-01],
[ 9.4100e-01],
[-1.4470e+00],
[-1.5692e+00],
[-2.3187e-01],
[ 1.2676e-01],
[ 4.1811e-01],
[ 5.5474e-01],
[-2.3442e+00],
[ 5.8383e-02],
[ 1.3507e+00],
[ 7.2094e-02],
[-1.8635e+00],
[-6.0708e-03],
[ 8.4099e-01],
[-4.6581e-01],
[ 9.0715e-02],
[-3.3725e-01],
[-6.5586e-02],
[-1.6582e+00],
[-1.7860e+00],
[-1.3451e+00],
[ 3.1415e-01],
[-1.8071e-01],
[-8.3670e-01],
[-7.3421e-02],
[-1.3120e+00],
[-3.5372e-01],
[-1.7302e-01],
[-6.9677e-01],
[ 4.8681e-02],
[ 5.6953e-02],
[ 7.4473e-01],
[-5.1874e-01],
[-3.2718e+00],
[-7.9775e-01],
[-1.4623e+00],
[-2.2119e-01],
[-5.8949e-01],
[-2.6437e-01],
[-3.7868e-01],
[-3.5251e-01],
[ 3.4737e-01],
[-3.4126e-01],
[-2.5543e-01],
[ 9.5274e-01]])
```

In [6]:

```
print( "y : ", y.grad)
```

```
y :  tensor([[-5.0359e-01],
       [-5.4955e+00],
       [-6.3276e+00],
       [-1.5391e+00],
       [ 1.8909e+00],
       [-1.9428e-01],
       [-1.5594e+00],
       [-1.1908e+00],
       [ 2.0805e-01],
       [ 1.2536e-01],
       [ 4.2350e+00],
       [-1.4492e+00],
       [ 2.0470e+00],
       [-2.0340e+00],
       [-4.0445e+00],
       [-7.1654e+00],
       [ 1.1789e-03],
       [-6.9952e-01],
       [ 4.6882e+00],
       [-7.3978e+00],
       [ 1.0876e+00],
       [-5.4611e+00],
       [ 1.1440e+00],
       [-3.5421e+00],
       [-3.8805e+00],
       [ 5.9145e-01],
       [-1.0356e+00],
       [-2.4182e+00],
       [-4.3287e+00],
       [ 1.1855e+00],
       [-4.9652e+00],
       [-1.9603e+00],
       [-1.3705e-01],
       [-6.5819e+00],
       [-1.3409e+00],
       [-1.0759e+00],
       [-3.8352e+00],
       [-6.5918e+00],
       [ 4.9417e+00],
       [ 1.2949e+00],
       [-4.3655e+00],
       [-3.3869e-02],
       [-1.7448e+00],
       [-2.7365e+00],
       [-5.6471e-03],
       [-1.4492e+00],
       [-2.3127e+00],
       [ 9.0197e-01],
       [ 2.3061e+00],
       [-3.2590e+00],
       [-5.6517e-01],
       [-4.3055e+00],
       [-6.8187e+00],
       [-1.5562e+00],
       [ 4.3578e+00],
       [ 2.9913e+00],
       [-3.2709e+00],
       [-4.8494e+00],
       [-6.5777e-01],
       [-4.8459e+00],
       [-5.1222e+00],
       [-3.1545e+00],
       [-7.9555e+00],
       [-2.3079e+00],
```

```
        [ 3.1980e+00],
        [-3.0513e+00],
        [ 4.2261e+00],
        [ 6.3740e-01],
        [-7.4774e+00],
        [-3.7476e+00],
        [-2.3074e+00],
        [-1.2390e+00],
        [ 1.2138e+00],
        [-6.6892e+00],
        [-3.6901e+00],
        [-4.7836e+00],
        [ 1.2775e+00],
        [-1.0194e+00],
        [-3.9872e+00],
        [-9.7562e-01],
        [-4.6250e+00],
        [-1.2316e+00],
        [-2.4279e+00],
        [-3.0602e+00],
        [ 4.5898e-01],
        [-2.4982e+00],
        [ 1.8484e+00],
        [-2.4444e+00],
        [-7.1057e+00],
        [-7.4393e+00],
        [-5.7969e+00],
        [-4.8342e-01],
        [-2.0376e+00],
        [-6.5589e+00],
        [ 2.8100e+00],
        [-2.6542e+00],
        [ 1.1024e+00],
        [-1.1678e+00],
        [-3.1190e+00],
        [ 2.6604e+00]])
```

In [7]:

```
print( "w1 : ", w1.grad)
```

```
w1 :  tensor([[ 3.3607, -0.2481,  0.8175, -1.9934,  0.7060,  0.3435,  0.0506,
-2.4668,
         -0.0808,  2.9497,  0.0749, -1.3073,  1.5510,  5.1425,  0.4083,  4.987
2,
          1.6239, -0.0107, -0.8858, -0.3937]])
```

In [8]:

```
print( "w2 : ", w2.grad)
```

```
w2 :  tensor([[102.7233],
        [ 97.4983],
        [ 86.5097],
        [105.5513],
        [ 97.9687],
        [ 92.4397],
        [ 90.4275],
        [ 86.6384],
        [ 97.9460],
        [ 73.4251],
        [ 99.2528],
        [ 95.9744],
        [ 82.7673],
        [116.7129],
        [ 85.9269],
        [101.5734],
```

```
        [ 99.8130],
        [ 84.6388],
        [101.2863],
        [ 92.7071]])
```

In [9]:
```
print( "b1 : ", b1.grad)
```

```
b1 :  tensor([[ 0.2557,  0.0302, -0.0846,  ..., -0.0430,  0.1223, -0.0290],
        [ 0.6673,  0.7031, -0.3876,  ..., -0.3169,  0.7183, -0.2683],
        [ 1.2436,  0.5236, -0.4475,  ..., -0.9619,  1.4208, -0.5739],
        ...,
        [ 0.4680,  0.1652, -0.2023,  ..., -0.1761,  0.1586, -0.1262],
        [ 0.9887,  0.4841, -0.5709,  ..., -0.4819,  0.7607, -0.3046],
        [-1.1204, -0.3583,  0.4860,  ...,  0.3617, -0.4004,  0.2100]])
```

In [10]:
```
print( "b2 : ", b2.grad)
```

```
b2 :  tensor([[ 5.0359e-01],
        [ 5.4955e+00],
        [ 6.3276e+00],
        [ 1.5391e+00],
        [-1.8909e+00],
        [ 1.9428e-01],
        [ 1.5594e+00],
        [ 1.1908e+00],
        [-2.0805e-01],
        [-1.2536e-01],
        [-4.2350e+00],
        [ 1.4492e+00],
        [-2.0470e+00],
        [ 2.0340e+00],
        [ 4.0445e+00],
        [ 7.1654e+00],
        [-1.1789e-03],
        [ 6.9952e-01],
        [-4.6882e+00],
        [ 7.3978e+00],
        [-1.0876e+00],
        [ 5.4611e+00],
        [-1.1440e+00],
        [ 3.5421e+00],
        [ 3.8805e+00],
        [-5.9145e-01],
        [ 1.0356e+00],
        [ 2.4182e+00],
        [ 4.3287e+00],
        [-1.1855e+00],
        [ 4.9652e+00],
        [ 1.9603e+00],
        [ 1.3705e-01],
        [ 6.5819e+00],
        [ 1.3409e+00],
        [ 1.0759e+00],
        [ 3.8352e+00],
        [ 6.5918e+00],
        [-4.9417e+00],
        [-1.2949e+00],
        [ 4.3655e+00],
        [ 3.3869e-02],
        [ 1.7448e+00],
        [ 2.7365e+00],
        [ 5.6471e-03],
        [ 1.4492e+00],
```

```
[ 2.3127e+00],
[-9.0197e-01],
[-2.3061e+00],
[ 3.2590e+00],
[ 5.6517e-01],
[ 4.3055e+00],
[ 6.8187e+00],
[ 1.5562e+00],
[-4.3578e+00],
[-2.9913e+00],
[ 3.2709e+00],
[ 4.8494e+00],
[ 6.5777e-01],
[ 4.8459e+00],
[ 5.1222e+00],
[ 3.1545e+00],
[ 7.9555e+00],
[ 2.3079e+00],
[-3.1980e+00],
[ 3.0513e+00],
[-4.2261e+00],
[-6.3740e-01],
[ 7.4774e+00],
[ 3.7476e+00],
[ 2.3074e+00],
[ 1.2390e+00],
[-1.2138e+00],
[ 6.6892e+00],
[ 3.6901e+00],
[ 4.7836e+00],
[-1.2775e+00],
[ 1.0194e+00],
[ 3.9872e+00],
[ 9.7562e-01],
[ 4.6250e+00],
[ 1.2316e+00],
[ 2.4279e+00],
[ 3.0602e+00],
[-4.5898e-01],
[ 2.4982e+00],
[-1.8484e+00],
[ 2.4444e+00],
[ 7.1057e+00],
[ 7.4393e+00],
[ 5.7969e+00],
[ 4.8342e-01],
[ 2.0376e+00],
[ 6.5589e+00],
[-2.8100e+00],
[ 2.6542e+00],
[-1.1024e+00],
[ 1.1678e+00],
[ 3.1190e+00],
[-2.6604e+00]])
```

手動求導

In [11]:

```
# print( "w1 : ", w1_grad)
# w1_grad
w1_grad = -2 * x.t().mm((y-yp).mm(w2.t()).mul(hs).mul(1-hs))
print( "w1 : ", w1_grad)
```

```
w1 :  tensor([[ 3.3607, -0.2481,  0.8175, -1.9934,  0.7060,  0.3435,  0.0506,
-2.4668,
         -0.0808,  2.9497,  0.0749, -1.3073,  1.5510,  5.1425,  0.4083,  4.987
2,
          1.6239, -0.0107, -0.8858, -0.3937]], grad_fn=<MulBackward0>)
```

In [12]:

```
# print( "w2 : ", w2_grad)
# w2_grad
w2_grad = -2 * hs.t().mm(y-yp)
print( "w2 : ", w2_grad)
```

```
w2 :  tensor([[102.7233],
        [ 97.4983],
        [ 86.5097],
        [105.5513],
        [ 97.9687],
        [ 92.4397],
        [ 90.4275],
        [ 86.6384],
        [ 97.9460],
        [ 73.4251],
        [ 99.2528],
        [ 95.9744],
        [ 82.7673],
        [116.7129],
        [ 85.9269],
        [101.5734],
        [ 99.8130],
        [ 84.6388],
        [101.2863],
        [ 92.7071]], grad_fn=<MulBackward0>)
```

In [13]:

```
# print( "b1 : ", b1_grad)
# b1_grad
b1_grad = -2 * ((y-yp).mm(w2.t()).mul(hs).mul((1-hs)))
print( "b1 : ", b1_grad)
```

```
b1 :  tensor([[ 0.2557,  0.0302, -0.0846,  ..., -0.0430,  0.1223, -0.0290],
        [ 0.6673,  0.7031, -0.3876,  ..., -0.3169,  0.7183, -0.2683],
        [ 1.2436,  0.5236, -0.4475,  ..., -0.9619,  1.4208, -0.5739],
        ...,
        [ 0.4680,  0.1652, -0.2023,  ..., -0.1761,  0.1586, -0.1262],
        [ 0.9887,  0.4841, -0.5709,  ..., -0.4819,  0.7607, -0.3046],
        [-1.1204, -0.3583,  0.4860,  ...,  0.3617, -0.4004,  0.2100]],
       grad_fn=<MulBackward0>)
```

In [14]:

```
# print( "b2 : ", b2_grad)
# b2_grad
b2_grad = -2 * (y - yp)
print( "b2 : ", b2_grad)
```

```
b2 :  tensor([[ 5.0359e-01],
        [ 5.4955e+00],
        [ 6.3276e+00],
        [ 1.5391e+00],
        [-1.8909e+00],
        [ 1.9428e-01],
        [ 1.5594e+00],
        [ 1.1908e+00],
        [-2.0805e-01],
        [-1.2536e-01],
        [-4.2350e+00],
```

```
[ 1.4492e+00],
[-2.0470e+00],
[ 2.0340e+00],
[ 4.0445e+00],
[ 7.1654e+00],
[-1.1789e-03],
[ 6.9952e-01],
[-4.6882e+00],
[ 7.3978e+00],
[-1.0876e+00],
[ 5.4611e+00],
[-1.1440e+00],
[ 3.5421e+00],
[ 3.8805e+00],
[-5.9145e-01],
[ 1.0356e+00],
[ 2.4182e+00],
[ 4.3287e+00],
[-1.1855e+00],
[ 4.9652e+00],
[ 1.9603e+00],
[ 1.3705e-01],
[ 6.5819e+00],
[ 1.3409e+00],
[ 1.0759e+00],
[ 3.8352e+00],
[ 6.5918e+00],
[-4.9417e+00],
[-1.2949e+00],
[ 4.3655e+00],
[ 3.3869e-02],
[ 1.7448e+00],
[ 2.7365e+00],
[ 5.6471e-03],
[ 1.4492e+00],
[ 2.3127e+00],
[-9.0197e-01],
[-2.3061e+00],
[ 3.2590e+00],
[ 5.6517e-01],
[ 4.3055e+00],
[ 6.8187e+00],
[ 1.5562e+00],
[-4.3578e+00],
[-2.9913e+00],
[ 3.2709e+00],
[ 4.8494e+00],
[ 6.5777e-01],
[ 4.8459e+00],
[ 5.1222e+00],
[ 3.1545e+00],
[ 7.9555e+00],
[ 2.3079e+00],
[-3.1980e+00],
[ 3.0513e+00],
[-4.2261e+00],
[-6.3740e-01],
[ 7.4774e+00],
[ 3.7476e+00],
[ 2.3074e+00],
[ 1.2390e+00],
[-1.2138e+00],
[ 6.6892e+00],
[ 3.6901e+00],
```

```
[ 4.7836e+00],
[-1.2775e+00],
[ 1.0194e+00],
[ 3.9872e+00],
[ 9.7562e-01],
[ 4.6250e+00],
[ 1.2316e+00],
[ 2.4279e+00],
[ 3.0602e+00],
[-4.5898e-01],
[ 2.4982e+00],
[-1.8484e+00],
[ 2.4444e+00],
[ 7.1057e+00],
[ 7.4393e+00],
[ 5.7969e+00],
[ 4.8342e-01],
[ 2.0376e+00],
[ 6.5589e+00],
[-2.8100e+00],
[ 2.6542e+00],
[-1.1024e+00],
[ 1.1678e+00],
[ 3.1190e+00],
[-2.6604e+00]], grad_fn=<MulBackward0>)
```

In [15]:
```
# Kan Horst
```

In [ ]:
```

```