$$F(A,B) = \frac{1}{N}\sum \ln\left(1+\exp\left(-y_n\left(A\cdot z_n + B\right)\right)\right)$$

1. $$\frac{\partial F(A,B)}{\partial A} = \frac{1}{N}\sum^{N} \frac{\exp(-y_n(Az_n+B))}{1+\exp(-y_n(Az_n+B))}\cdot(-y_n z_n) = -\frac{1}{N}\sum_{n=1}^{N} p_n y_n z_n$$

$$\frac{\partial F(A,B)}{\partial B} = -\frac{1}{N}\sum_{n=1}^{N} p_n y_n$$

$$\nabla F(A,B) = -\frac{1}{N}\sum_{n=1}^{N} y_n p_n (z_n, 1)$$

2. $$H(F) = \begin{pmatrix} \frac{\partial^2 F}{\partial A^2} & \frac{\partial^2 F}{\partial A \partial B} \\ \frac{\partial^2 F}{\partial B \partial A} & \frac{\partial^2 F}{\partial B^2} \end{pmatrix}$$

$$\frac{\partial^2 F(A,B)}{\partial A^2} = \frac{\partial}{\partial A}\left(-\frac{1}{N}\sum p_n y_n z_n\right) = \frac{1}{N}\sum_{n=1}^{N} z_n^2 p_n(1-p_n)$$

$$\frac{\partial^2 F(A,B)}{\partial A \partial B} = \frac{1}{N}\sum^{N} z_n p_n(1-p_n)$$

$$\frac{\partial^2 F(A,B)}{\partial B^2} = \frac{1}{N}\sum p_n(1-p_n)$$

$$H(F) = \frac{1}{N}\sum_{n=1}^{N} p_n(1-p_n)\begin{bmatrix} z_n^2 & z_n \\ z_n & 1 \end{bmatrix}$$

3. when $\gamma \to \infty$ $\quad K(x, x') = \exp(-\gamma \|x-x'\|^2) \to 0$
$(x \neq x')$

$\Rightarrow$ $b = 1$ or -1. $\quad \alpha = 0$.

Double A

4. two samples $\begin{cases} (x_1, & x_1 - x_1^2) \\ (x_2, & \underbrace{x_2 - x_2^2}_{y}) \end{cases}$

$$E_{OLS} = (w_1 x_1 + w_0 - x_1 + x_1^2)^2 + (w_1 x_2 + w_0 - x_2 + x_2^2)^2$$

$$\frac{\partial E_{OLS}}{\partial w_1} = 2(w_1 x_1 + w_0 - x_1 + x_1^2) x_1 + 2(w_1 x_0 + w_0 - x_2 + x_2^2) x_2 = 0$$

$$\Rightarrow \quad w_1 = (1 - x_1 - x_2)$$

$$\frac{\partial E_{OLS}}{\partial w_0} = 2(w_1 x_1 + w_0 - x_1 + x_1^2) + 2(w_1 x_2 + w_0 - x_2 + x_2^2) = 0$$

$$\Rightarrow \quad w_0 = +x_1 x_2$$

$$\Rightarrow \quad g(x) = (1 - x_1 - x_2) x + x_1 \cdot x_2$$

$$\bar{g}(x) = E[g(x)] = \frac{1}{4}$$

5. $$u_n (y_n - w^T x_n)^2 = (\sqrt{u_n} y_n - w^T \sqrt{u_n} x_n)^2$$

$$\{(\tilde{x}_n, \hat{y}_n)\}_{n=1}^{N} = \{(\sqrt{u_n} x_n, \sqrt{u_n} y_n)\}_{n=1}^{N}$$

6. error rate $= 0.22 = \varepsilon_t$.

$$\frac{u_+^{(2)}}{u_-^{(2)}} = \frac{\varepsilon_t}{1 - \varepsilon_t} = \frac{22}{78} = \frac{11}{39}$$

7. First, all positive and negative are two decision stumps (D.S)

For each dimension, there are $2 \times M$ interval between $-M$ and M, each could be $+1$ or -1

total d dimensions, $\Rightarrow$ total $4dM+2$ DSs.

8. $K_{ds}(x, x')$ is inner product of $\phi_{ds}(x)$

$$= 4dM+2 - 2 \cdot \sum_{n=1}^{|\mathcal{S}|} |x_n - x_n'|$$

if $x = x'$ $\Rightarrow$ $K(x, x') = 4dM+2$.

Double A

In [164]:
```
import pandas as pd
import numpy as np
from sklearn.linear_model import Ridge
from sklearn.utils import resample
```

In [165]:
```
df = np.loadtxt('hw2_lssvm_all.dat')
```

In [166]:
```
dftrain = df[0:400]
dftest = df[400:]
```

In [167]:
```
Lambda = [0.05, 0.5, 5, 50, 500]
```

In [ ]:
```
```

In [168]:
```
Ein = []
Eout = []
for l in Lambda:
    clf = Ridge(alpha=l)
    clf.fit(dftrain[:,0:-1], dftrain[:,-1])
    y_pred_in = (clf.predict(dftrain[:,0:-1])>0)*2-1
    y_pred_out = (clf.predict(dftest[:,0:-1])>0)*2-1
    Ein = Ein + [sum(y_pred_in != dftrain[:,-1])/len(y_pred_in)]
    Eout = Eout + [sum(y_pred_out != dftest[:,-1])/len(y_pred_out)]
```

In [169]:
```
Ein
```

Out[169]:
```
[0.3175, 0.3175, 0.3175, 0.32, 0.3225]
```

In [170]:
```
Eout
```

Out[170]:
```
[0.36, 0.36, 0.36, 0.37, 0.37]
```

With λ increases, Ein and Eout increase.

In [171]:
```
Ein = []
Eout = []
for l in Lambda:
    y_bag_out = np.zeros(100)
    y_bag_in = np.zeros(400)
    for t in range(250):
        df_bs = resample(dftrain,n_samples=400,random_state=t)
        clf = Ridge(alpha=l)
        clf.fit(df_bs[:,0:-1], df_bs[:,-1])
        y_bag_out = y_bag_out + (( clf.predict(dftest[:,0:-1]) > 0 )*2-1)
        y_bag_in = y_bag_in +   (( clf.predict(dftrain[:,0:-1]) > 0 )*2-1)



    Ein = Ein + [sum( ((y_bag_in>0)*2-1) != dftrain[:,-1] ) / len(y_bag_in)]
    Eout = Eout + [sum( ((y_bag_out>0)*2-1) != dftest[:,-1] )/ len(y_bag_out)]
```

In [172]:
```
Ein
```

Out[172]:
```
[0.3175, 0.3175, 0.3175, 0.3175, 0.3225]
```

In [173]:
```
Eout
```

Out[173]:
```
[0.36, 0.36, 0.36, 0.36, 0.37]
```

With λ increases, Ein and Eout increase. And there is no improvement after using bootstrpping.

In [ ]:
```
```

In [ ]:
```
```

In [ ]:
```
```