HW 5

Patrick Gardocki

2023-11-06

2. SVM

1.

$max\ \frac{2c}{||w||}\ \ subject\ to\ y^i(w^Tx^i+b)\geq c \equiv max\ \frac{1}{||w||}\ \ subject\ to\ y^i(w^Tx^i+b)\geq 1$

The two statements are equivalent. Given any solution, the other statement can be achieved by scaling w and b. This does not change the goodness of the classifiers.

2.

$L(w,b,\alpha)=\frac{1}{2}w^Tw+\sum_{i=1}^m \alpha_i(1-y_i(w^Tx_i+b))$

$\frac{\partial L(w,b,\alpha)}{\partial w}=w-\sum_{i=1}^m \alpha_i y_i x_i=0 \therefore w=\sum_{i=1}^m \alpha_i y_i x_i$

This implies that the weight vector is a linear combination of the feature vectors.

3.

Given KKT conditions: $\alpha_i(1-y_i(w^Tx_i+b))=0$;
$1-y_i(w^Tx_i+b)=0\ if\ x_i\ is\ on\ margin$
$else\ 1-y_i(w^Tx_i+b)<0$

$\therefore \alpha_i=0$ if x_i not on margin.

The sum from part 2, $w=\sum_{i=1}^m \alpha_i y_i x_i$ will only contain data points on the margin.

4.

4.a

The training points will remain linearly separable if x_3 remains on the left side of the line formed by positive samples, x_1 and x_2. Therefore, for $0\leq h\leq 1$, the training points will remain linearly separable.

4.b

The orientation of the maximum margin decision boundary does not change as h changes in the linearly separable range. The 2 hyper planes that defined the boundary are parallel to each other when the data points are separable.

3. Neural Network and Backward propogation

1.

$l(w, \alpha, \beta) = \sum_{i=1}^{m} (y_i - \sigma(w^T z_i))^2$

$\frac{\partial l(w,\alpha,\beta)}{\partial w} = \sum_{i=1}^{m} \frac{\partial}{\partial w}(y_i - \sigma(w^T z_i))^2 = \sum_{i=1}^{m} -2(y_i - \sigma(w^T z_i)) * \frac{\partial}{\partial w}(\sigma u^i)$

Given: $\sigma(x) = \frac{1}{1+e^{-x}}$: $\frac{\partial \sigma}{\partial x} = \frac{-1}{(1+e^{-x})} \frac{e^{-x}}{1+e^{-x}} = \sigma(x)(\frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}) = \sigma(x)(1 - \sigma(x)$

$\frac{\partial}{\partial w}(\sigma(w^T z^i)) = \sigma(w^T z^i)(1 - \sigma(w^T z^i))z^i$

$\therefore \frac{\partial l(w,\alpha,\beta)}{\partial w} = -\sum_{i=1}^{m} 2(y_i - \sigma(u_i))\sigma(u_i)(1 - \sigma(u_i))z_i$

2.

$\frac{\partial l(w,\alpha,\beta)}{\partial \alpha} = \frac{\partial l}{\partial z_1^i} \frac{\partial z_1^i}{\partial \alpha} = (-\sum_{i=1}^{m} 2(y_i - \sigma(u_i))\sigma(u_i)(1 - \sigma(u_i))w_1)(\sigma(\alpha^T x^i)(1 - \sigma(\alpha^T x^i))x^i)$

$\frac{\partial l(w,\alpha,\beta)}{\partial \beta} = \frac{\partial l}{\partial z_2^i} \frac{\partial z_2^i}{\partial \beta} = (-\sum_{i=1}^{m} 2(y_i - \sigma(u_i))\sigma(u_i)(1 - \sigma(u_i))w_2)(\sigma(\beta^T x^i)(1 - \sigma(\beta^T x^i))x^i)$

4. Feature Selection and Change-point detection

1.

$I(U,C) = \frac{N_{11}}{N} log_2 \frac{NN_{11}}{N_{1.}N_{.1}} + \frac{N_{01}}{N} log_2 \frac{NN_{01}}{N_{0.}N_{.1}} + \frac{N_{10}}{N} log_2 \frac{NN_{10}}{N_{1.}N_{.0}} + \frac{N_{00}}{N} log_2 \frac{NN_{00}}{N_{0.}N_{.0}}$

$I(prize, spam) = \frac{150}{16160} log_2 \frac{150*16160}{160*1150} + \frac{1000}{16160} log_2 \frac{1000*16160}{16000*1150} + \frac{10}{16160} log_2 \frac{10*16160}{160*15010} + \frac{15000}{16160} log_2 \frac{15000*16160}{16000*15010} = 0.03302$

$I(hello, spam) = \frac{145}{16160} log_2 \frac{145*16160}{160*11145} + \frac{11000}{16160} log_2 \frac{11000*16160}{16000*11145} + \frac{15}{16160} log_2 \frac{15*16160}{160*5015} + \frac{5000}{16160} log_2 \frac{5000*16160}{16000*5015} = 0.001948$

Given $I(prize, spam) > I(hello, spam)$, 'prize' is more informative for deciding email spam.

2.

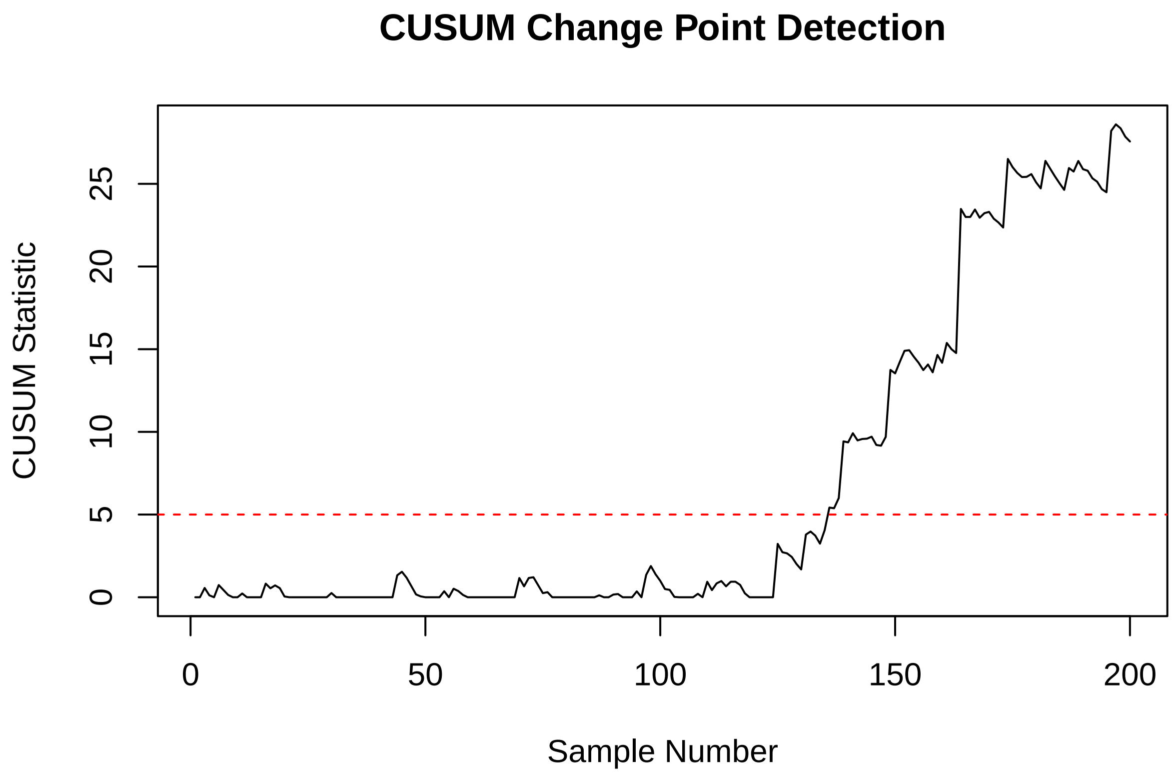
CUSUM Statistic Derivation

$L_t = \log\left(\frac{f_1(x_t)}{f_0(x_t)}\right) = \log\left(\frac{N(0.5,1.5)}{N(0,1)}\right) = \log\left(\frac{\frac{1}{\sqrt{2\pi 1.5^2}} \exp\left(-\frac{(x_t-0.5)^2}{2(1.5^2)}\right)}{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_t^2}{2}\right)}\right)$

Simplifying this expression:

$L_t = \log\left(\frac{1}{\sqrt{1.5^2}} \exp\left(-\frac{(x_t-0.5)^2}{2(1.5^2)} + \frac{x_t^2}{2}\right)\right)$

$W_t = max(W_{t-1} + L_t, 0)$




```
## [1] 27.55964
```

5. Medical Imaging Reconstuction

1.

2.