CSE 546: Homework 3

Due on November ***, 2016

Brian de Silva

0 Collaborators and Acknowledgements

1 PCA and reconstruction

1.1 Matrix Algebra Review

1. Recall that for $A \in \mathbb{R}^{n\times d}$ and $C \in \mathbb{R}^{d\times n}$, $(AC)_{ij} = \sum_{k=1}^{d} a_{ik}c_{kj}$. Also, $(B^T)_{ij} = B_{ji}$. Hence

$$(AB^T)_{ij} = \sum_{i=1}^{d}(A)_{ik}(B^T)_{jk} = \sum_{i=1}^{d} a_{ik}b_{kj}.$$

Plugging this into the definition of the trace gives

$$\begin{aligned}\mathrm{Tr}(AB^T) &= \sum_{i=1}^{n}\left(\sum_{k=1}^{d}(AB^T)_{ii}\right)\\ &= \sum_{i=1}^{n}\left(\sum_{k=1}^{d} a_{ik}b_{ik}\right).\end{aligned}$$

Similarly,

$$(B^TA)_{ij} = \sum_{k=1}^{n} b_{ki}a_{kj}$$

and, by switching the order of addition,

$$\begin{aligned}\mathrm{Tr}(B^TA) &= \sum_{i=1}^{d}\left(B^TA\right)_{ii}\\ &= \sum_{i=1}^{d}\left(\sum_{k=1}^{n} b_{ki}a_{ki}\right)\\ &= \sum_{i=1}^{n}\left(\sum_{k=1}^{d} b_{ki}a_{ki}\right)\\ &= \mathrm{Tr}(AB^T).\end{aligned}$$

2. The outer equality follows from the definition of the trace:

$$\begin{aligned}\mathrm{Tr}(\Sigma) = \mathrm{Tr}\left(\tfrac{1}{n}X^TX\right) = \frac{1}{n}\sum_{i=1}^{d}(X^TX)_{ii} &= \frac{1}{n}\sum_{i=1}^{d}\left(\sum_{k=1}^{n} x_{ik}^2\right)\\ &= \frac{1}{n}\sum_{i=1}^{d}\|X_i\|^2.\end{aligned}$$

For the other equality, we need some standard linear algebra results. Since Σ is symmetric and real, it has a real orthogonal eigendecomposition. That is to say, there exists an orthogonal matrix Q and a diagonal matrix Λ with diagonal entries $\lambda_1, \lambda_2, \dots, \lambda_d$ such that

$$\Sigma = Q\Lambda Q^T.$$

Using our result from part 1, we have

$$\mathrm{Tr}(\Sigma) = \mathrm{Tr}\left(Q\Lambda Q^T\right) = \mathrm{Tr}\left((Q\Lambda)Q^T\right) = \mathrm{Tr}\left(Q^TQ\Lambda\right) = \mathrm{Tr}(\Lambda) = \sum_{i=1}^{d}\lambda_i.$$

2 SVMs: Hinge loss and mistake bounds

1. To show that $\ell((x,y),w) = \max\{0, 1-w\cdot x\}$ is convex with respect to w, we will need two observations. First, for any $a,k\in\mathbb{R}$ with $k\geq 0$, we have that

$$\max\{0, ka\} = k\max\{0,a\}.$$

In the case $ka<0$, both sides of the equality are 0. If $ka\geq 0$, then $a\geq 0$ and the equality still holds. Next, for any $a,b\in\mathbb{R}$, we have

$$\max\{0, a+b\} \leq \max\{0,a\} + \max\{0,b\}.$$

If both a and b are negative, then the above is an equality. If exactly one of a and b is negative and $a+b\leq 0$ then the inequality clearly holds. If both are nonnegative then it is also an equality.

Recall that a function f is convex if for any $t\in[0,1]$ and for any x,y in its domain, $f(tx+(1-t)y)\leq tf(x)+(1-t)f(y)$. We will show that $\ell((x,y),w)$ has this property with respect to w. Let $w_1,w_2\in\mathbb{R}^d$ be arbitrary and let $t\in[0,1]$. Then $0\leq(1-t)\leq 1$, and so

$$\begin{aligned}
\ell\left((x,y), tw_1+(1-t)w_2\right) &= \max\{0, 1-y(tw_1+(1-t)w_2)\cdot x\}\\
&= \max\{0, 1-tyw_1\cdot x-(1-t)yw_2\cdot x\}\\
&= \max\{0, 1+(t-t)-tyw_1\cdot x-(1-t)yw_2\cdot x\}\\
&= \max\{0, [t-tyw_1\cdot x]+[(1-t)-(1-t)yw_2\cdot x]\}\\
&\leq \max\{0, t-tyw_1\cdot x\} + \max\{(1-t)-(1-t)`yw_2\cdot x\}\\
&= t\max\{0, 1-yw_1\cdot x\} + (1-t)\max\{1-yw_2\cdot x\}\\
&= t\ell((x,y),w_1) + (1-t)\ell((x,y),w_2).
\end{aligned}$$

Therefore $\ell((x,y),w)$ is convex with respect to w.

2. By its definition it is clear that $0\leq\ell((x,y),w)$. If $y_i=\text{sgn}(w\cdot x_i)$ then $y_iw\cdot x_i\geq 0$, implying that $1-y_iw\cdot x_i\leq 1$. Hence $\ell((x_i,y_i),w)=\max\{0,1-y_iw\cdot x_i\}\leq 1$. Combining these bounds we get

$$0\leq\ell((x_i,y_i),w)\leq 1$$

for correctly classified points.

3. Observe that if we misclassify a point (so that $y_i=-\text{sgn}(w\cdot x_i)$) then $y_iw\cdot x_i\geq 0$. Hence $\ell((x_i,y_i),w)\geq 1$ for misclassified points. Let $I\subset\{1,2,\dots,n\}$ be the indices corresponding to the data points which w misclassifies. It follows that $|I|=M(w)$. By the previous part we know that for correct classifications the hinge loss is bounded between 0 and 1. Putting this all together we obtain

$$M(w) = \sum_{i=1}^{M(w)} 1 = \sum_{i\in I} 1 \leq \sum_{i\in I}\ell((x_i,y_i),w) \leq \sum_{i=1}^{n}\ell((x_i,y_i),w) = \sum_{i=1}^{n}\max\{0, 1-y_iw\cdot x_i\}.$$

Dividing both sides by n gives the desired result.