

COMPUTER SYSTEMS RESEARCH

Code Writeup of your program, example report form 2009-2010

1. Your name: _Sam Rush_____, Period: _4_
2. Date of this version of your program: _4/8/10_____
3. Project title: _Machine Learning of the College Admissions Process_____
4. Describe how your program runs as of this version. Include
 - files that may be needed
 - algorithms, specific procedures or methods you wrote
 - kinds of input your program uses
 - screenshots, what kinds of output does your program have
 - does your program handle errors, or does it crash on errors of input?
 - tests: summarize the basic analysis and testing of this version of your program

Due to the massive size (over 30 pages) of my second quarter code writeup, I have only included new code in this document.

The program currently takes the numerical ID of a college in my database as input and outputs the number of correct and incorrect predictions that it makes for the class of 2010's college admissions. It accomplishes this using a linear least squares using QR decomposition, which involves the Gram-Schmidt process $O(n^3)$ and Gaussian elimination $O(n^3)$. The program does not currently handle errors of input, but the only possible error is to give a college with fewer than three applicants, so that should not generally be a problem. The program predicts admissions for most colleges at a rate above 70% when trained on a random subset of applications. Below is a screenshot for the predictions for Brown University.

Brown University

Array ([2] => -3.55794696164 [1] => 6.12914378926 [0] => -1.94717005387)

Prediction Actual

```
0 1
1 1
1 0
0 0
1 1
1 0
0 0
0 1
0 0
0 1
0 0
0 0
0 1
0 0
0 0
0 0
1 1
0 0
0 0
0 1
0 1
0 0
1 1
0 0
0 0
0 0
-0 0
0 0
```

17 / 25

The following files are new to this quarter:

qr.php

This is the main file for the quarter. It contains everything that is necessary to the linear algebra that I am doing in my project. This includes methods to decompose a matrix into its QR components, solve a consistent system $Ax=B$, multiply matrices, transpose a matrix. In addition, it also contains the code to take in information from the database, manipulate it, and make predictions for student decisions.

```
<?php
```

```
$outcomes=array("Waitlisted"=>0.5,"Accepted"=>1,"Attending"=>1,"Rejected"=>0);
```

```
$genders=array("M"=>-1,"F"=>1);
```

```
$people=array();
```

```
$weights=array();
```

```
$decisions=array();
```

```
$cid=$_GET['cid'];
```

```
mysql_connect('localhost','Physics','tjphysics');
```

```
mysql_select_db('destinations');
```

```
list($college)=mysql_fetch_row(mysql_query("SELECT name FROM colleges WHERE  
code='$cid';"));
```

```
echo "<h4>$college</h4>\n";
```

```
$query="SELECT user.SAT24C, user.GPA, user.Gender, applications.Decision FROM applications  
JOIN user ON (applications.User = user.ID)
```

```
WHERE applications.Code='$cid' AND user.SAT24C IS NOT NULL AND user.GPA IS NOT NULL  
AND user.Gender IS NOT NULL AND applications.Decision IS NOT NULL;";
```

```
$result=mysql_query($query);
```

```
while(list($sat,$gpa,$gender,$decision)=mysql_fetch_row($result)){
```

```
$people[]=array($sat/2400.0,$gpa/4.5,1);
```

```
$decisions[]=array($outcomes[$decision]);
```

```
$weights[]=1;
```

```
}
```

```
$b=array(
```

```
array(12,-51,4),
```

```
array(6,167,-68),
```

```
array(-4,24,-41),
```

```
);
```

```
$d=array(
```

```
array(1,0,0),
```

```

array(0,1,0),
array(0,0,1)
);
$N=count($people[0]);
$A=R($people);
$extra=count($A);
for($i=$N;$i<$extra;$i++)
unset($A[$i]);
#print_r($decisions);
$B=matmult(transpose(Q($people)),$decisions);
$C=array();
for($i=0;$i<count($B);$i++)
$C[]=$B[$i][0];
#matprint($B);
#matprint($A);
$weights=matsolve($A,$C);
print_r($weights);
echo "<table border=0>
<tr><td>Prediction</td><td>Actual</td></tr>\n";
$correct=0;
for($i=0;$i<count($people);$i++){
$score=0;
for($j=0;$j<count($weights);$j++)
$score+=$weights[$j]*$people[$i][$j];
$score=round($score);
echo "<tr><td>$score</td><td>".$decisions[$i][0]."</td></tr>";
if($score==$decisions[$i][0])$correct++;
}
echo "</table> $correct / ".count($people);
#matprint(Q($people));
#matprint($B);
#matprint($decisions);
#for($i=$N;$i<$extra;$i++)
#{

```

```

#unset($B[$i]);
#unset($decisions[0][$i]);
#}

#$c=Q($b);
#matprint(matmult($b,$d));
function matprint($c){
foreach($c as $arr)
{
foreach($arr as $element)
echo "$element ";
echo "\n";
}
}
function matmult($a,$b)
{
$c=array();
for($i=0;$i<count($a);$i++)
{
$c[]=$array();
}
for($i=0;$i<count($a);$i++){
for($j=0;$j<count($a[0]);$j++)
for($k=0;$k<count($b[0]);$k++)
$c[$i][$k]+=$a[$i][$j]*$b[$j][$k];
}
return $c;
}
function transpose($a){
$aT=array();
for($i=0;$i<count($a[0]);$i++)
$aT[]=$array();

for($i=0;$i<count($a);$i++)

```

```

for($j=0;$j<count($a[0]);$j++)
$aT[$j][$i]=$a[$i][$j];

return $aT;
}
function R($a){
$c=matmult(transpose(Q($a)),$a);
for($i=0;$i<count($c[0]);$i++)
for($j=0;$j<$i;$j++)
$c[$i][$j]=0;
return $c;
}
function Q($a){//Q part of the Q-R factorization of x using the Gram-Schmidt Orthogonalization
process
$k=count($a[0]);//cols
$n=count($a);//rows
#for($j=0;$j<$k;$j++)
#for($i=0;$i<$n;$i++)
#$a[$i][$j]=$x[$i][$j]+0.0;
for($j=0;$j<$k;$j++)
{
for($i=0;$i<$j;$i++){
$d=0;
for($l=0;$l<$n;$l++)
$d+=$a[$l][$i]*$a[$l][$j];
for($l=0;$l<$n;$l++)
$a[$l][$j] -= $d*$a[$l][$i];
}
$norm=0;
for($l=0;$l<$n;$l++)
$norm+=pow($a[$l][$j],2);
$invnorm=1.0/sqrt($norm);
for($l=0;$l<$n;$l++)
$a[$l][$j]*=$invnorm;
}

```

```

}
return $a;
}
function matsolve($a,$b){
$n=count($b);
for($j=0;$j<$n;$j++) {
$max=$j;
for($i=$j+1;$i<$n;$i++)
if(abs($a[$i][$j])>abs($a[$max][$j]))$max=$i;

$temp=$a[$j];$a[$j]=$a[$max];$a[$max]=$temp;
$t=$b[$j];$b[$j]=$b[$max];$b[$max]=$t;

for($i=$j+1;$i<$n;$i++){
$r=$a[$i][$j]/$a[$j][$j];
$b[$i]=$b[$i]-$r*$b[$j];
for($k=$j;$k<$n;$k++)
$a[$i][$k] -= $r*$a[$j][$k];
}
}

$x=array();
for($i=$n-1;$i>=0;$i--){
$sum=0.0;
for($j=$i+1;$j<$n;$j++)
$sum+=$a[$i][$j]*$x[$j];
$x[$i]=($b[$i]-$sum)/$a[$i][$i];
}
return $x;
}

```

In addition, I have made some updates to the Senior Destinations website in order to garner more participation (it's worked!). The new files are below.

recover.php

This is a file which allows for password recovery. People are apparently incapable of remembering their passwords. The addition of this files ensures that everyone can use the site.

```
<?php
session_start();
include("../libraries/lib.php");
makeheader("Password Recovery");
if(isset($_POST[tjid]))
{
$tjid=$_POST[tjid];
mysql_connect('localhost','Physics',$dbpass);
mysql_select_db('destinations');
$query="SELECT First, Last FROM user WHERE TJID='$tjid'";
$result=mysql_query($query);
if(mysql_num_rows($result)==0)
echo "Error: User does not exist.";
else{
$query2="SELECT First, Last FROM user WHERE TJID='$tjid' AND
TIMESTAMPDIFF(day,LastRecovery,NOW())>0";
$result2=mysql_query($query2);
if(mysql_num_rows($result2)==0)
echo "Error: You can only reset your password once in any 24 hour time period.";
else{
list($first,$last)=mysql_fetch_row($result);
$pass=substr(sha1(sha1($tjid.rand(1,1000000)).rand(1,1000000)),0,10);
$hash=sha1($pass);
$query = "UPDATE user SET Backup='$hash', LastRecovery=NOW() WHERE TJID = '$tjid'";
$result = mysql_query($query) or die ("Error in query: $query. " . mysql_error());
#echo mysql_num_rows($result);

// free result set memory
$email=$tjid."@tjhsst.edu";
$headers = 'From: "Senior Destinations" <tjseniordestinations@gmail.com>' . "\r\n";
$headers .= 'To: '.$first.' '.$last.' <'.$email.'>';
```

```
$subject='Senior Destinations Password Recovery';
$message="Your temporary password is $pass. If you did not request this email, just ignore it; your
old password will still work.";
if(mail(",$subject,$message,$headers)){

echo "Recovery email has been sent.";
}
else echo "Error in mail server, please contact site administrator.";
}
}
}
else{
?>
<script type="text/javascript" src="SHA1.js"></script>
    <form name="login_form" action="<?php echo $_SERVER['PHP_SELF']?>" method="post">
        <table id="login_box" nowrap>
            <tr>
                <td>TJ ID <font size="-2">(e.g. 2010flast)</font>:</td>
                <td><input name="tjid" type="text" size="15" style="background-color:#FFFFFF"
/></td>
            </tr>
            <tr>
                <td align="right" colspan="2"><input type="submit" value="Submit"
/></td>
            </tr>
        </table>
    </form>
    <script language="javascript" type="text/javascript">
        document.login_form.tjid.focus();
    </script>
<?php
}
makefooter();
?>
```


stats.php

This provides an easy way for me to monitor the participation of the site.

```
<?php
```

```
include('libraries/lib.php');
```

```
makeheader('Senior Destinations - Site Statistics');
```

```
mysql_connect('localhost','Physics',$dbpass);
```

```
mysql_select_db('destinations');
```

```
echo "<table border=0>";
```

```
list($registered)=mysql_fetch_row(mysql_query("SELECT COUNT(ID) FROM user;"));
```

```
echo "<tr><td>Registered Users:</td>
```

```
<td>$registered</td></tr>\n";
```

```
list($activated)=mysql_fetch_row(mysql_query("SELECT COUNT(ID) FROM user WHERE  
Activated=1;"));
```

```
echo "<tr><td>Activated Users:</td>
```

```
<td>$activated</td></tr>\n";
```

```
echo "<tr><td></td></tr>";
```

```
list($registeredS)=mysql_fetch_row(mysql_query("SELECT COUNT(ID) FROM user WHERE  
Grade=12;"));
```

```
echo "<tr><td>Registered Seniors:</td>
```

```
<td>$registeredS</td></tr>\n";
```

```
list($activatedS)=mysql_fetch_row(mysql_query("SELECT COUNT(ID) FROM user WHERE  
Activated=1 AND Grade=12;"));
```

```
echo "<tr><td>Activated Seniors:</td>
```

```
<td>$activatedS</td></tr>\n";
```

```
echo "<tr><td></td></tr>";
```

```
list($appsubmitters)=mysql_fetch_row(mysql_query("SELECT COUNT(DISTINCT user) FROM  
applications;"));
```

```
echo "<tr><td>Users w/App:</td>
```

```
<td>$appsubmitters</td></tr>\n";
```

```
list($decsubmitters)=mysql_fetch_row(mysql_query("SELECT COUNT(DISTINCT user) FROM  
applications WHERE Decision IS NOT NULL OR Waitlisted=1 OR Deferred=1;"));
```

```
echo "<tr><td>Users w/Decision:</td>
<td>$decsubmitters</td></tr>\n";
```

```
echo "<tr><td></td></tr>";
```

```
list($apps)=mysql_fetch_row(mysql_query("SELECT COUNT(user) FROM applications;"));
```

```
echo "<tr><td>Applications:</td>
<td>$apps</td></tr>\n";
```

```
list($decs)=mysql_fetch_row(mysql_query("SELECT COUNT(user) FROM applications WHERE
Decision IS NOT NULL OR Waitlisted=1 OR Deferred=1;"));
```

```
echo "<tr><td>Decisions:</td>
<td>$decs</td></tr>\n";
```

```
list($dests)=mysql_fetch_row(mysql_query("SELECT COUNT(DISTINCT user) FROM applications
WHERE Decision='Attending;"));
```

```
echo "<tr><td>Destinations:</td>
<td>$dests</td></tr>\n";
```

```
echo "</table>";
```

```
makefooter();
```

```
?>
```

recent.php

This file provides a way for me to see what people are adding to the site as it happens. This is important because I often have to delete obviously fake content which could interfere with the success of this project. In addition, there is an rss feed that I have written, but the code is nearly identical to that of this file, so I have not included it.

```
<?php
```

```
include('libraries/lib.php');
```

```
makeheader('Senior Destinations - Site Statistics');
```

```
mysql_connect('localhost','Physics',$dbpass);
```

```
mysql_select_db('destinations');
```

```
echo "<table border=0>";
```

```
list($registered)=mysql_fetch_row(mysql_query("SELECT COUNT(ID) FROM user;"));
```

```
echo "<tr><td>Registered Users:</td>
<td>$registered</td></tr>\n";
```

```
list($activated)=mysql_fetch_row(mysql_query("SELECT COUNT(ID) FROM user WHERE
Activated=1;"));
```

```
echo "<tr><td>Activated Users:</td>
<td>$activated</td></tr>\n";
```

```
echo "<tr><td></td></tr>";
```

```
list($registeredS)=mysql_fetch_row(mysql_query("SELECT COUNT(ID) FROM user WHERE
Grade=12;"));
```

```
echo "<tr><td>Registered Seniors:</td>
<td>$registeredS</td></tr>\n";
```

```
list($activatedS)=mysql_fetch_row(mysql_query("SELECT COUNT(ID) FROM user WHERE
Activated=1 AND Grade=12;"));
```

```
echo "<tr><td>Activated Seniors:</td>
<td>$activatedS</td></tr>\n";
```

```
echo "<tr><td></td></tr>";
```

```
list($appsubmitters)=mysql_fetch_row(mysql_query("SELECT COUNT(DISTINCT user) FROM
applications;"));
```

```
echo "<tr><td>Users w/App:</td>
<td>$appsubmitters</td></tr>\n";
```

```
list($decsubmitters)=mysql_fetch_row(mysql_query("SELECT COUNT(DISTINCT user) FROM
applications WHERE Decision IS NOT NULL OR Waitlisted=1 OR Deferred=1;"));
```

```
echo "<tr><td>Users w/Decision:</td>
<td>$decsubmitters</td></tr>\n";
```

```
echo "<tr><td></td></tr>";
```

```
list($apps)=mysql_fetch_row(mysql_query("SELECT COUNT(user) FROM applications;"));
```

```
echo "<tr><td>Applications:</td>
<td>$apps</td></tr>\n";
```

```
list($decs)=mysql_fetch_row(mysql_query("SELECT COUNT(user) FROM applications WHERE
Decision IS NOT NULL OR Waitlisted=1 OR Deferred=1;"));
```

```
echo "<tr><td>Decisions:</td>
<td>$decs</td></tr>\n";
```

```
list($dests)=mysql_fetch_row(mysql_query("SELECT COUNT(DISTINCT user) FROM applications
WHERE Decision='Attending'"));
echo "<tr><td>Destinations:</td>
<td>$dests</td></tr>\n";

echo "</table>";
makefooter();
?>
```

5. What do you expect to work on next quarter, in relation to the goal of your project for the year?

Next quarter, I will finish the prediction part of the project by adding in the machine learning of nonlinear least squares, which is a small extension upon this method but will bring a massive improvement and additional algorithmic complexity. After that, I intend to investigate the program's decision about how to count factors such as race and gender into the algorithm.