Big Data lab

Class exercise 7 - Hackathon

SQL

```
INSERT   # insert records into the database
SELECT   # query the database
sqldf()  # perform queries in R
table()  # enumerate a contingency table / cross tabulation / pivot
order()   # sort the data - return the index order
nrow()    # number of rows
ncol()    # number of columns
```

reminder : install packages into R

```
install.packages('package-name')

installed.packages() # retrieve all insatalled packages

sessionInfo()     # session information (+attached packages)
```

The data

- **BX-Users**
  Contains the users. Note that user IDs (`User-ID`) have been anonymized and map to integers. Demographic data is provided (`Location`, `Age`) if available. Otherwise, these fields contain NULL-values.

- **BX-Books**
  Books are identified by their respective ISBN. Invalid ISBNs have already been removed from the dataset. Moreover, some content-based information is given (`Book-Title`, `Book-Author`, `Year-Of-Publication`, `Publisher`), obtained from Amazon Web Services. Note that in case of several authors, only the first is provided. URLs linking to cover images are also given, appearing in three different flavours (`Image-URL-S`, `Image-URL-M`, `Image-URL-L`), i.e., small, medium, large. These URLs point to the Amazon web site.

- **BX-Book-Ratings**
  Contains the book rating information. Ratings (`Book-Rating`) are either explicit, expressed on a scale from 1-10 (higher values denoting higher appreciation), or implicit, expressed by 0.

SQL file

```
-- MySQL dump 9.11
--
-- Host: localhost    Database: Book-Crossing
-- ------------------------------------------------------
-- Server version       4.0.20a-debug

--
-- Table structure for table `BX-Books`
--

CREATE TABLE `BX-Books` (
  `ISBN` varchar(13) binary NOT NULL default '',
  `Book-Title` varchar(255) default NULL,
  `Book-Author` varchar(255) default NULL,
  `Year-Of-Publication` int(10) unsigned default NULL,
  `Publisher` varchar(255) default NULL,
  `Image-URL-S` varchar(255) binary default NULL,
  `Image-URL-M` varchar(255) binary default NULL,
  `Image-URL-L` varchar(255) binary default NULL,
  PRIMARY KEY  (`ISBN`)
) TYPE=MyISAM;

--
-- Dumping data for table `BX-Books`
--

INSERT INTO `BX-Books` VALUES ('0195153448','Classical Mythology','Mark P. O. Morford',2002,'Oxford University Press','http://images.amazon.com/images/P/0195153448.01.THUMBZZZ.
INSERT INTO `BX-Books` VALUES ('0002005018','Clara Callan','Richard Bruce Wright',2001,'HarperFlamingo Canada','http://images.amazon.com/images/P/0002005018.01.THUMBZZZ.jpg','h
INSERT INTO `BX-Books` VALUES ('0060973129','Decision in Normandy','Carlo D\'Este',1991,'HarperPerennial','http://images.amazon.com/images/P/0060973129.01.THUMBZZZ.jpg','http:/
INSERT INTO `BX-Books` VALUES ('0374157065','Flu: The Story of the Great Influenza Pandemic of 1918 and the Search for the Virus That Caused It','Gina Bari Kolata',1999,'Farrar
INSERT INTO `BX-Books` VALUES ('0393045218','The Mummies of Urumchi','E. J. W. Barber',1999,'W. W. Norton &amp;amp; Company','http://images.amazon.com/images/P/0393045218.01.THUMBZ
INSERT INTO `BX-Books` VALUES ('0399135782','The Kitchen God\'s Wife','Amy Tan',1991,'Putnam Pub Group','http://images.amazon.com/images/P/0399135782.01.THUMBZZZ.jpg','http://i
INSERT INTO `BX-Books` VALUES ('0425176428','What If?: The World\'s Foremost Military Historians Imagine What Might Have Been','Robert Cowley',2000,'Berkley Publishing Group','
INSERT INTO `BX-Books` VALUES ('0671870432','PLEADING GUILTY','Scott Turow',1993,'Audioworks','http://images.amazon.com/images/P/0671870432.01.THUMBZZZ.jpg','http://images.amaz
INSERT INTO `BX-Books` VALUES ('0679425608','Under the Black Flag: The Romance and the Reality of Life Among the Pirates','David Cordingly',1996,'Random House','http://images.a
INSERT INTO `BX-Books` VALUES ('074322678X','Where You\'ll Find Me: And Other Stories','Ann Beattie',2002,'Scribner','http://images.amazon.com/images/P/074322678X.01.THUMBZZZ.j
INSERT INTO `BX-Books` VALUES ('0771074670','Nights Below Station Street','David Adams Richards',1988,'Emblem Editions','http://images.amazon.com/images/P/0771074670.01.THUMBZZ
INSERT INTO `BX-Books` VALUES ('080652121X','Hitler\'s Secret Bankers: The Myth of Swiss Neutrality During the Holocaust','Adam Lebor',2000,'Citadel Press','http://images.amazo
INSERT INTO `BX-Books` VALUES ('0887841740','The Middle Stories','Sheila Heti',2004,'House of Anansi Press','http://images.amazon.com/images/P/0887841740.01.THUMBZZZ.jpg','http
INSERT INTO `BX-Books` VALUES ('1552041778','Jane Doe','R. J. Kaiser',1999,'Mira Books','http://images.amazon.com/images/P/1552041778.01.THUMBZZZ.jpg','http://images.amazon.com
INSERT INTO `BX-Books` VALUES ('1558746218','A Second Chicken Soup for the Woman\'s Soul (Chicken Soup for the Soul Series)','Jack Canfield',1998,'Health Communications','http:
INSERT INTO `BX-Books` VALUES ('1567407781','The Witchfinder (Amos Walker Mystery Series)','Loren D. Estleman',1998,'Brilliance Audio - Trade','http://images.amazon.com/images/
INSERT INTO `BX-Books` VALUES ('1575663937','More Cunning Than Man: A Social History of Rats and Man','Robert Hendrickson',1999,'Kensington Publishing Corp.','http://images.ama
INSERT INTO `BX-Books` VALUES ('1881320189','Goodbye to the Buttermilk Sky','Julia Oliver',1994,'River City Pub','http://images.amazon.com/images/P/1881320189.01.THUMBZZZ.jpg',
INSERT INTO `BX-Books` VALUES ('0440234743','The Testament','John Grisham',1999,'Dell','http://images.amazon.com/images/P/0440234743.01.THUMBZZZ.jpg','http://images.amazon.com/
INSERT INTO `BX-Books` VALUES ('0452264464','Beloved (Plume Contemporary Fiction)','Toni Morrison',1994,'Plume','http://images.amazon.com/images/P/0452264464.01.THUMBZZZ.jpg','
```

Import MySQL into sqlite3

- Create SQLite3 tables (can convert the syntax in existing schemas)

  use `CREATE TABLE`

- Optional: grep the `INSERT` statements to a new sql file

```
grep “INSERT” dump.sql > MySQLdump_filtered.sql
```

- Optional: split `INSERT` statements into transactions
- convert \’ to “ (quote-quote)
- run sqlite3 with the new file

```
sqlite3 empty_sqlite_db.db < MySQLdump_filtered.sql
```

Import MySQL into sqlite3

- Create SQLite3 tables (can convert the syntax in existing schemas)

  use `CREATE TABLE`
- Optional: grep the `INSERT` statements to a new sql file

  ```
  grep "INSERT" dump.sql > MySQLdump_filtered.sql
  ```
- Optional: split `INSERT` statements into transactions
- convert \' to '' (quote-quote)
- run sqlite3 with the new file

  ```
  sqlite3 empty_sqlite_db.db < MySQLdump_filtered.sql
  ```

Import MySQL into sqlite3

- Create SQLite3 tables (car
  use **CREATE TABLE**
- Optional: grep the **INSERT**
  grep "INSERT" dump.s … SQLdump_filtered.sql
- Optional: split **INSERT** statements into transactions
- convert \' to '' (quote-quot
- run sqlite3 with the new fil
  sqlite3 empty_sqlite …

```
BEGIN TRANSACTION;
INSERT INTO `BX-Books` VALUES ('0195153448','Classical Mythology','Mark P. O. Morford',2002,'Oxford University Press','http://images.amazon.com/images/P/0195153
INSERT INTO `BX-Books` VALUES ('0002005018','Clara Callan','Richard Bruce Wright',2001,'HarperFlamingo Canada','http://images.amazon.com/images/P/0002005018.01.
INSERT INTO `BX-Books` VALUES ('0060973129','Decision in Normandy','Carlo D\'Este',1991,'HarperPerennial','http://images.amazon.com/images/P/0060973129.01.THUMB
INSERT INTO `BX-Books` VALUES ('0374157065','Flu: The Story of the Great Influenza Pandemic of 1918 and the Search for the Virus That Caused It','Gina Bari Kola
INSERT INTO `BX-Books` VALUES ('0393045218','The Mummies of Urumchi','E. J. W. Barber',1999,'W. W. Norton &amp;amp; Company','http://images.amazon.com/images/P/0393
INSERT INTO `BX-Books` VALUES ('0399135782','The Kitchen God\'s Wife','Amy Tan',1991,'Putnam Pub Group','http://images.amazon.com/images/P/0399135782.01.THUMBZZ
INSERT INTO `BX-Books` VALUES ('0425176428','What If?: The World\'s Foremost Military Historians Imagine What Might Have Been','Robert Cowley',2000,'Berkley Pub
INSERT INTO `BX-Books` VALUES ('0671870432','PLEADING GUILTY','Scott Turow',1993,'Audioworks','http://images.amazon.com/images/P/0671870432.01.THUMBZZZ.jpg','ht
INSERT INTO `BX-Books` VALUES ('0679425608','Under the Black Flag: The Romance and the Reality of Life Among the Pirates','David Cordingly',1996,'Random House',
INSERT INTO `BX-Books` VALUES ('074322678X','Where You\'ll Find Me: And Other Stories','Ann Beattie',2002,'Scribner','http://images.amazon.com/images/P/07432267
INSERT INTO `BX-Books` VALUES ('0771074670','Nights Below Station Street','David Adams Richards',1988,'Emblem Editions','http://images.amazon.com/images/P/07710
INSERT INTO `BX-Books` VALUES ('080652121X','Hitler\'s Secret Bankers: The Myth of Swiss Neutrality During the Holocaust','Adam Lebor',2000,'Citadel Press','htt
INSERT INTO `BX-Books` VALUES ('0887841740','The Middle Stories','Sheila Heti',2004,'House of Anansi Press','http://images.amazon.com/images/P/0887841740.01.THU
INSERT INTO `BX-Books` VALUES ('1552041778','Jane Doe','R. J. Kaiser',1999,'Mira Books','http://images.amazon.com/images/P/1552041778.01.THUMBZZZ.jpg','http://i
INSERT INTO `BX-Books` VALUES ('1558746218','A Second Chicken Soup for the Woman\'s Soul (Chicken Soup for the Soul Series)','Jack Canfield',1998,'Health Commun
INSERT INTO `BX-Books` VALUES ('1567407781','The Witchfinder (Amos Walker Mystery Series)','Loren D. Estleman',1998,'Brilliance Audio - Trade','http://images.am
INSERT INTO `BX-Books` VALUES ('1575663937','More Cunning Than Man: A Social History of Rats and Man','Robert Hendrickson',1999,'Kensington Publishing Corp.','h
INSERT INTO `BX-Books` VALUES ('1881320189','Goodbye to the Buttermilk Sky','Julia Oliver',1994,'River City Pub','http://images.amazon.com/images/P/1881320189.0
INSERT INTO `BX-Books` VALUES ('0440234743','The Testament','John Grisham',1999,'Dell','http://images.amazon.com/images/P/0440234743.01.THUMBZZZ.jpg','http://im
INSERT INTO `BX-Books` VALUES ('0452264464','Beloved (Plume Contemporary Fiction)','Toni Morrison',1994,'Plume','http://images.amazon.com/images/P/0452264464.01
END TRANSACTION;
BEGIN TRANSACTION;
INSERT INTO `BX-Books` VALUES ('0971880107','Wild Animus','Rich Shapero',2004,'Too Far','http://images.amazon.com/images/P/0971880107.01.THUMBZZZ.jpg','http://i
INSERT INTO `BX-Books` VALUES ('0345402871','Airframe','Michael Crichton',1997,'Ballantine Books','http://images.amazon.com/images/P/0345402871.01.THUMBZZZ.jpg',
INSERT INTO `BX-Books` VALUES ('0345417623','Timeline','MICHAEL CRICHTON',2000,'Ballantine Books','http://images.amazon.com/images/P/0345417623.01.THUMBZZZ.jpg',
                                                                                                                                . Lewis',1996,'Scribner','http://images.amazon.com/images/P/0684823802.01.THUMBZZZ.jpg
                                                                                                                                LIPS',2003,'Random House Trade Paperbacks','http://images.amazon.com/images/P/0375759
INSERT INTO `BX-Books` VALUES ('0425163091','Chocolate Jesus','Stephan Jaramillo',1998,'Berkley Publishing Group','http://images.amazon.com/images/P/0425163091.0
INSERT INTO `BX-Books` VALUES ('3404921038','Wie Barney es sieht.','Mordecai Richler',2002,'L<C3>?¼bbe','http://images.amazon.com/images/P/3404921038.01.THUMBZZ
INSERT INTO `BX-Books` VALUES ('3442353866','Der Fluch der Kaiserin. Ein Richter- Di- Roman.','Eleanor Cooney',2001,'Goldmann','http://images.amazon.com/images/
INSERT INTO `BX-Books` VALUES ('3442410665','Sturmzeit. Roman.','Charlotte Link',1991,'Goldmann','http://images.amazon.com/images/P/3442410665.01.THUMBZZZ.jpg',
INSERT INTO `BX-Books` VALUES ('3442446937','Tage der Unschuld.','Richard North Patterson',2000,'Goldmann','http://images.amazon.com/images/P/3442446937.01.THUM
INSERT INTO `BX-Books` VALUES ('0375406328','Lying Awake','Mark Salzman',2000,'Alfred A. Knopf','http://images.amazon.com/images/P/0375406328.01.THUMBZZZ.jpg','
INSERT INTO `BX-Books` VALUES ('0446310786','To Kill a Mockingbird','Harper Lee',1988,'Little Brown &amp;amp; Company','http://images.amazon.com/images/P/0446310786
INSERT INTO `BX-Books` VALUES ('0449005615','Seabiscuit: An American Legend','LAURA HILLENBRAND',2002,'Ballantine Books','http://images.amazon.com/images/P/0449
INSERT INTO `BX-Books` VALUES ('0060168013','Pigs in Heaven','Barbara Kingsolver',1993,'Harpercollins','http://images.amazon.com/images/P/0060168013.01.THUMBZZZ
INSERT INTO `BX-Books` VALUES ('038078243X','Miss Zukas and the Raven\'s Dance','Jo Dereske',1996,'Avon','http://images.amazon.com/images/P/038078243X.01.THUMBZ
INSERT INTO `BX-Books` VALUES ('055321215X','Pride and Prejudice','Jane Austen',1983,'Bantam','http://images.amazon.com/images/P/055321215X.01.THUMBZZZ.jpg','ht
INSERT INTO `BX-Books` VALUES ('067176537X','The Therapeutic Touch: How to Use Your Hands to Help or to Heal','Dolores Krieger',1979,'Fireside','http://images.a
INSERT INTO `BX-Books` VALUES ('0061099686','Downtown','Anne Rivers Siddons',1995,'HarperTorch','http://images.amazon.com/images/P/0061099686.01.THUMBZZZ.jpg','
INSERT INTO `BX-Books` VALUES ('0553582909','Icebound','Dean R. Koontz',2000,'Bantam Books','http://images.amazon.com/images/P/0553582909.01.THUMBZZZ.jpg','http
INSERT INTO `BX-Books` VALUES ('0671888587','I\'ll Be Seeing You','Mary Higgins Clark',1994,'Pocket','http://images.amazon.com/images/P/0671888587.01.THUMBZZZ.j
END TRANSACTION;
BEGIN TRANSACTION;
```

Import MySQL into sqlite3

- Create SQLite3 tables (can convert the syntax in existing schemas)

  use `CREATE TABLE`
- Optional: grep the `INSERT` statements to a new sql file

  ```
  grep "INSERT" dump.sql > MySQLdump_filtered.sql
  ```
- Optional: split `INSERT` statements into transactions
- convert `\'` to `''` (quote-quote)
- run sqlite3 with the new file

  ```
  sqlite3 empty_sqlite_db.db < MySQLdump_filtered.sql
  ```

Import MySQL into sqlite3

- Create SQLite3 tables (can convert the syntax in existing schemas)
  use `CREATE TABLE`
- Optional: grep the `INSERT` statements to a new sql file

```
grep "INSERT" dump.sql > MySQLdump_filtered.sql
```

- Optional: split `INSERT` statements into transactions
- convert `\'` to `''` (quote-quote)
- run sqlite3 with the new file

```
sqlite3 empty_sqlite_db.db < MySQLdump_filtered.sql
```

'Miss Zukas and the Raven\'s Dance'

'Miss Zukas and the Raven''s Dance'

```
sed -i.bak -e "s/\\\'/''/g" BX-Books.sql
```

Import MySQL into sqlite3

- Create SQLite3 tables (can convert the syntax in existing schemas)

  use `CREATE TABLE`
- Optional: grep the `INSERT` statements to a new sql file

  ```
  grep "INSERT" dump.sql > MySQLdump_filtered.sql
  ```
- Optional: split `INSERT` statements into transactions
- convert \' to '' (quote-quote)
- run sqlite3 with the new file

  ```
  sqlite3 empty_sqlite_db.db < MySQLdump_filtered.sql
  ```

Exercise
Recommendation system

- Load SQL data
- Convert to your favorite database
  - https://guillempg.wordpress.com/2014/12/05/importing-a-mysql-dump-into-sqlite3
  - https://blog.jetbrains.com/webide/2012/11/sql-support-and-database-tools/
  - https://stackoverflow.com/questions/5324601/mysql-to-oracle
- collect statistics
  - how many users?
  - how many books?
  - how many ratings? histograms of ratings (by user/by book)
  - top-10 books with ratings? histogram of ratings
  - top-10 users that rated? histogram of ratings

basics.txt

```
# Team: <team name>
# Date: <date>
# Database name   <name>

3.a) how many users?      <number>
3.b) how many books?      <number>
3.c) how many ratings?    <number>
3.d) histogram of user-ratings <table(num ratings, num users)>
     (how many users have rated N times? <number>)
+-----+-------+
| bin | N     |
+-----+-------+
|   1 | XXXXX |
|   2 | XXXXX |
|   3 | XXXXX |
|   4 | XXXXX |
|   5 | XXXXX |
|   6 | XXXXX |
|   7 | XXXXX |
|   8 | XXXXX |
...
3.e) histogram of book-ratings <table(num ratings, num users)>
     (how many books have been rated N times? <number>)
+-----+-------+
| bin | N     |
+-----+-------+
|   1 | XXXXX |
|   2 | XXXXX |
|   3 | XXXXX |
|   4 | XXXXX |
|   5 | XXXXX |
|   6 | XXXXX |
|   7 | XXXXX |
|   8 | XXXXX |
...
```

```
3.f) top-10 rated books?    <table(name,num ratings)>
+----------+-------+
| name     | N     |
+----------+-------+
| Catch 22 | XXXXX |
...
3.g) top-10 active users?   <table(name, num ratings)>
+-------+-------+
| name  | N     |
+-------+-------+
|   226 | XXXXX |
...
```

Exercise

Recommendation system

- Create the
  - Similarity matrix
  - user-item rating matrix
- train a model with holdout (k-fold)
- Prediction:
  - get user recommendations for 500 books
  - get new book ISBN
  - give recommendation

Clean the data

Confusion Matrix

		Predicted condition			
	Total population = P + N	Predicted condition **positive** (PP)	Predicted condition **negative** (PN)	Informedness, bookmaker informedness (BM) = TPR + TNR − 1	Prevalence threshold (PT) = $\frac{\sqrt{TPR \cdot FPR} - FPR}{TPR - FPR}$
Actual condition	Actual condition positive (P)	**True positive (TP), hit**	**False negative (FN), Type II error, miss, overestimation**	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power = $\frac{TP}{P}$ = 1−FNR	False negative rate (FNR), miss rate = $\frac{FN}{P}$ = 1−TPR
Actual condition	Actual condition negative (N)	**False positive (FP), Type I error, false alarm, underestimation**	**True negative (TN), correct rejection**	False positive rate (FPR), probability of false alarm, fall-out = $\frac{FP}{N}$ = 1−TNR	True negative rate (TNR), specificity (SPC), selectivity = $\frac{TN}{N}$ = 1−FPR
	Prevalence = $\frac{P}{P+N}$	Positive predictive value (PPV), precision = $\frac{TP}{PP}$ = 1−FDR	False omission rate (FOR) = $\frac{FN}{PN}$ = 1−NPV	Positive likelihood ratio (LR+) = $\frac{TPR}{FPR}$	Negative likelihood ratio (LR−) = $\frac{FNR}{TNR}$
	Accuracy (ACC) = $\frac{TP+TN}{P+N}$	False discovery rate (FDR) = $\frac{FP}{PP}$ = 1−PPV	Negative predictive value (NPV) = $\frac{TN}{PN}$ = 1−FOR	Markedness (MK), deltaP (Δp) = PPV + NPV − 1	Diagnostic odds ratio (DOR) = $\frac{LR+}{LR-}$
	Balanced accuracy (BA) = $\frac{TPR+TNR}{2}$	F_1 score = $\frac{2 \cdot PPV \cdot TPR}{PPV+TPR}$ = $\frac{2TP}{2TP+FP+FN}$	Fowlkes–Mallows index (FM) = $\sqrt{PPV \cdot TPR}$	Threat score (TS), critical success index (CSI) = $\frac{TP}{TP+FN+FP}$	Matthews correlation coefficient (MCC) = $\sqrt{TPR \cdot TNR \cdot PPV \cdot NPV} - \sqrt{FNR \cdot FPR \cdot FOR \cdot FDR}$

Sources: [13][14][15][16][17][18][19][20]

view · talk · edit

Confusion Matrix

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

		Predicted condition		Sources: [13][14][15][16][17][18][19][20]	view · talk · edit
	Total population = P + N	Predicted condition **positive** (PP)	Predicted condition **negative** (PN)	Informedness, bookmaker informedness (BM) = TPR + TNR − 1	Prevalence threshold (PT) = $\frac{\sqrt{\text{TPR}\cdot\text{FPR}} - \text{FPR}}{\text{TPR} - \text{FPR}}$
Actual condition	Actual condition positive (P)	**True positive (TP), hit**	**False negative (FN), Type II error, miss, overestimation**	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power = $\frac{\text{TP}}{\text{P}}$ = 1 − FNR	False negative rate (FNR), miss rate = $\frac{\text{FN}}{\text{P}}$ = 1 − TPR
	Actual condition negative (N)	**False positive (FP), Type I error, false alarm, underestimation**	**True negative (TN), correct rejection**	False positive rate (FPR), probability of false alarm, fall-out = $\frac{\text{FP}}{\text{N}}$ = 1 − TNR	True negative rate (TNR), specificity (SPC), selectivity = $\frac{\text{TN}}{\text{N}}$ = 1 − FPR
	Prevalence = $\frac{\text{P}}{\text{P} + \text{N}}$	Positive predictive value (PPV), precision = $\frac{\text{TP}}{\text{PP}}$ = 1 − FDR	False omission rate (FOR) = $\frac{\text{FN}}{\text{PN}}$ = 1 − NPV	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Negative likelihood ratio (LR−) = $\frac{\text{FNR}}{\text{TNR}}$
	Accuracy (ACC) = $\frac{\text{TP} + \text{TN}}{\text{P} + \text{N}}$	False discovery rate (FDR) = $\frac{\text{FP}}{\text{PP}}$ = 1 − PPV	Negative predictive value (NPV) = $\frac{\text{TN}}{\text{PN}}$ = 1 − FOR	Markedness (MK), deltaP (Δp) = PPV + NPV − 1	Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR−}}$
	Balanced accuracy (BA) = $\frac{\text{TPR} + \text{TNR}}{2}$	F_1 score = $\frac{2\cdot\text{PPV}\cdot\text{TPR}}{\text{PPV} + \text{TPR}}$ = $\frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$	Fowlkes–Mallows index (FM) = $\sqrt{\text{PPV}\cdot\text{TPR}}$	Threat score (TS), critical success index (CSI) = $\frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}}$	Matthews correlation coefficient (MCC) = $\frac{\sqrt{\text{TPR}\cdot\text{TNR}\cdot\text{PPV}\cdot\text{NPV}} - \sqrt{\text{FNR}\cdot\text{FPR}\cdot\text{FOR}\cdot\text{FDR}}}{}$

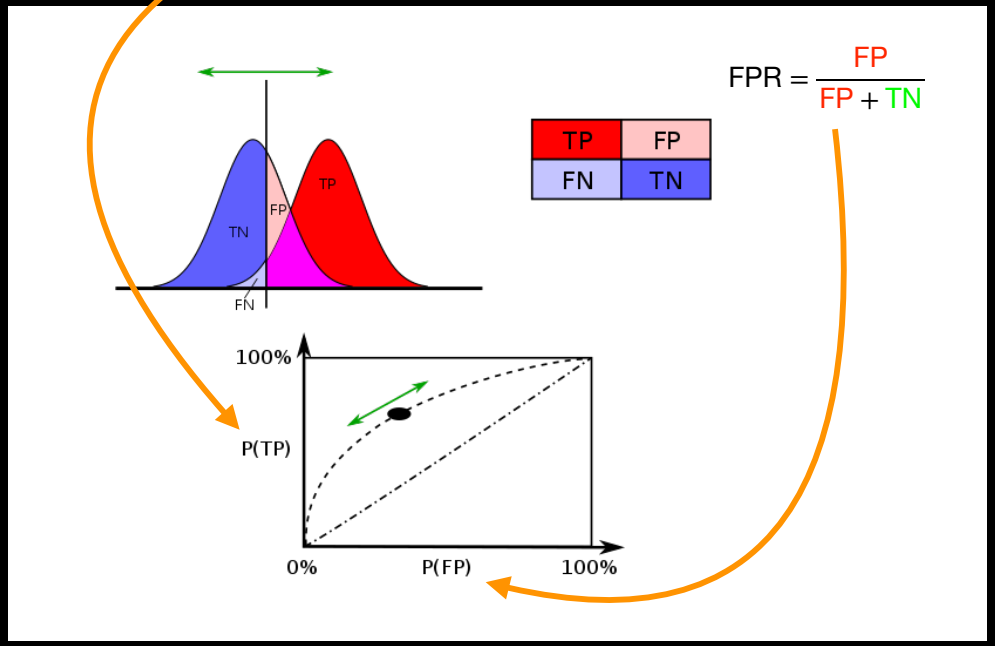
Confusion Matrix

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

		Predicted condition		Sources: [13][14][15][16][17][18][19][20]	view · talk · edit
	Total population = P + N	Predicted condition **positive** (PP)	Predicted condition **negative** (PN)	Informedness, bookmaker informedness (BM) = TPR + TNR − 1	Prevalence threshold (PT) = $\frac{\sqrt{\text{TPR}\cdot\text{FPR}} - \text{FPR}}{\text{TPR} - \text{FPR}}$
Actual condition	Actual condition positive (P)	**True positive (TP), hit**	**False negative (FN), Type II error, miss, overestimation**	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power = $\frac{\text{TP}}{\text{P}}$ = 1 − FNR	False negative rate (FNR), miss rate = $\frac{\text{FN}}{\text{P}}$ = 1 − TPR
Actual condition	Actual condition negative (N)	**False positive (FP), Type I error, false alarm, underestimation**	**True negative (TN), correct rejection**	False positive rate (FPR), probability of false alarm, fall-out = $\frac{\text{FP}}{\text{N}}$ = 1 − TNR	True negative rate (TNR), specificity (SPC), selectivity = $\frac{\text{TN}}{\text{N}}$ = 1 − FPR
	Prevalence = $\frac{\text{P}}{\text{P} + \text{N}}$	Positive predictive value (PPV), precision = $\frac{\text{TP}}{\text{PP}}$ = 1 − FDR	False omission rate (FOR) = $\frac{\text{FN}}{\text{PN}}$ = 1 − NPV	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Negative likelihood ratio (LR−) = $\frac{\text{FNR}}{\text{TNR}}$
	Accuracy (ACC) = $\frac{\text{TP} + \text{TN}}{\text{P} + \text{N}}$	False discovery rate (FDR) = $\frac{\text{FP}}{\text{PP}}$ = 1 − PPV	Negative predictive value (NPV) = $\frac{\text{TN}}{\text{PN}}$ = 1 − FOR	Markedness (MK), deltaP (Δp) = PPV + NPV − 1	Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR−}}$
	Balanced accuracy (BA) = $\frac{\text{TPR} + \text{TNR}}{2}$	F_1 score = $\frac{2\cdot\text{PPV}\cdot\text{TPR}}{\text{PPV} + \text{TPR}}$ = $\frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$	Fowlkes–Mallows index (FM) = $\sqrt{\text{PPV}\cdot\text{TPR}}$	Threat score (TS), critical success index (CSI) = $\frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}}$	Matthews correlation coefficient (MCC) = $\frac{\sqrt{\text{TPR}\cdot\text{TNR}\cdot\text{PPV}\cdot\text{NPV}} - \sqrt{\text{FNR}\cdot\text{FPR}\cdot\text{FOR}\cdot\text{FDR}}}{}$

Receiver Operating Characteristic

$$TPR = \frac{TP}{TP + FN}$$

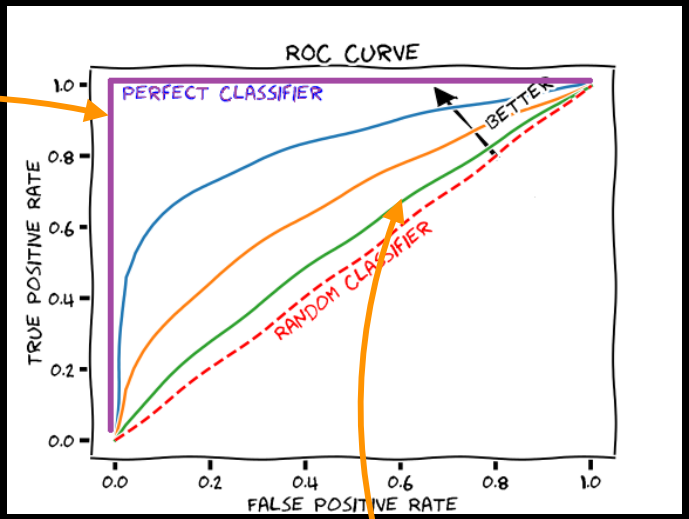


$$FPR = \frac{FP}{FP + TN}$$

		Predicted condition		Sources: [13][14][15][16][17][18][19][20]	view · talk · edit
	Total population = P + N	Predicted condition **positive** (PP)	Predicted condition **negative** (PN)	Informedness, bookmaker informedness (BM) = TPR + TNR − 1	Prevalence threshold (PT) = $\frac{\sqrt{TPR \cdot FPR} - FPR}{TPR - FPR}$
Actual condition	Actual condition positive (P)	**True positive (TP), hit**	**False negative (FN), Type II error, miss, overestimation**	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power = $\frac{TP}{P}$ = 1−FNR	False negative rate (FNR), miss rate = $\frac{FN}{P}$ = 1−TPR
	Actual condition negative (N)	**False positive (FP), Type I error, false alarm, underestimation**	**True negative (TN), correct rejection**	False positive rate (FPR), probability of false alarm, fall-out = $\frac{FP}{N}$ = 1−TNR	True negative rate (TNR), specificity (SPC), selectivity = $\frac{TN}{N}$ = 1−FPR
	Prevalence = $\frac{P}{P+N}$	Positive predictive value (PPV), precision = $\frac{TP}{PP}$ = 1−FDR	False omission rate (FOR) = $\frac{FN}{PN}$ = 1−NPV	Positive likelihood ratio (LR+) = $\frac{TPR}{FPR}$	Negative likelihood ratio (LR−) = $\frac{FNR}{TNR}$
	Accuracy (ACC) = $\frac{TP+TN}{P+N}$	False discovery rate (FDR) = $\frac{FP}{PP}$ = 1−PPV	Negative predictive value (NPV) = $\frac{TN}{PN}$ = 1−FOR	Markedness (MK), deltaP (Δp) = PPV + NPV − 1	Diagnostic odds ratio (DOR) = $\frac{LR+}{LR-}$
	Balanced accuracy (BA) = $\frac{TPR+TNR}{2}$	F_1 score = $\frac{2 \cdot PPV \cdot TPR}{PPV + TPR}$ = $\frac{2TP}{2TP + FP + FN}$	Fowlkes–Mallows index (FM) = $\sqrt{PPV \cdot TPR}$	Threat score (TS), critical success index (CSI) = $\frac{TP}{TP + FN + FP}$	Matthews correlation coefficient (MCC) = $\frac{\sqrt{TPR \cdot TNR \cdot PPV \cdot NPV} - \sqrt{FNR \cdot FPR \cdot FOR \cdot FDR}}{}$

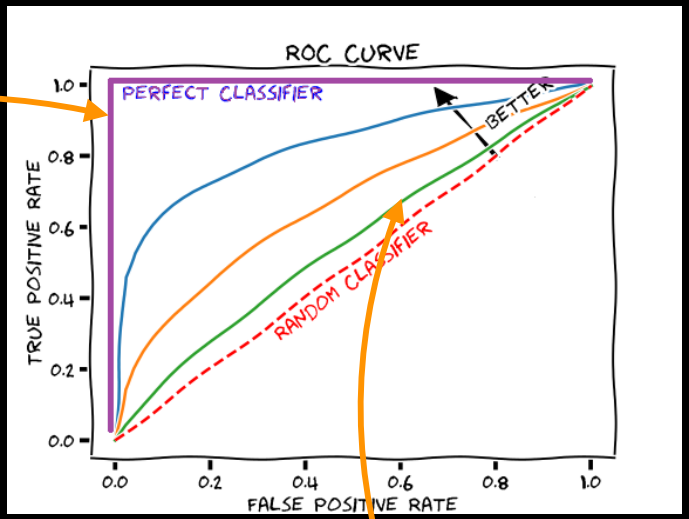
Receiver Operating Characteristic

a perfect model



a useless model

Receiver Operating Characteristic



a perfect model

random:

coin toss, fair coin ($p = \frac{1}{4}$)

blind guess on Heads

	H	T	
H	5	0	5
T	15	0	15
	20	0	20

a useless model

random:

coin toss, fair coin ($p = \frac{1}{2}$)

num successes = num failures

	H	T	
H	5	5	10
T	5	5	10
	10	10	20

$\text{TPR} = 0.5$

$\text{FPR} = 0.5$

random:

coin toss, fair coin ($p = \frac{1}{4}$)

	H	T	
H	2	3	5
T	8	7	15
	10	10	20

$\text{TPR} = 0.5$

$\text{FPR} = 0.53$