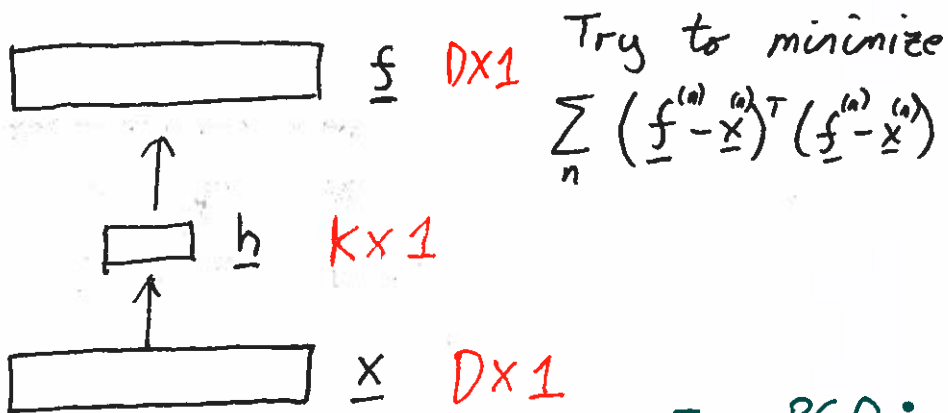


Dimensionality reduction, Auto-encoders



For PCA:

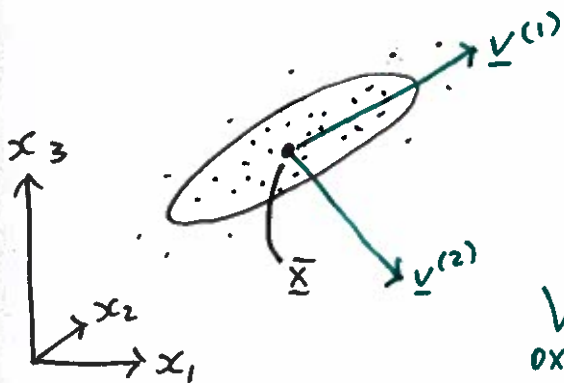
$$\underline{h} = g^{(1)}(W^{(1)}\underline{x} + \underline{b}^{(1)}) = V^T(\underline{x} - \bar{\underline{x}})$$

$$\underline{f} = g^{(2)}(W^{(2)}\underline{h} + \underline{b}^{(2)}) = V\underline{h} + \bar{\underline{x}}$$

Use $g^{(1)}$ to reduce dimensionality new data

$g^{(2)}$ synthesize high-dim objects from a few numbers, \underline{h}

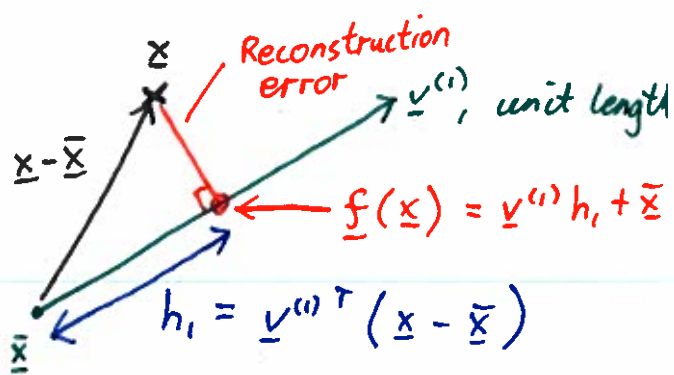
PCA Principal Components Analysis



$\underline{v}^{(k)}$ k^{th} eigenvector of $\text{cov}[X]$
(Tutorial 2, Q4b)

$$V_{D \times K} = \begin{bmatrix} | & | & & | \\ \underline{v}^{(1)} & \underline{v}^{(2)} & \dots & \underline{v}^{(K)} \\ | & | & & | \end{bmatrix}$$

When $k=1$:



Pythagoras:

$$\| \underline{x} - \bar{\underline{x}} \|^2 = \| \underline{h} \|^2 + \text{square error}$$

{ Data variance
{ Variance
+ square error

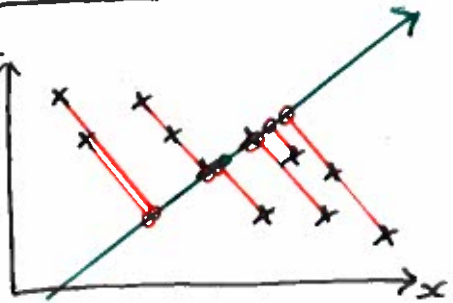
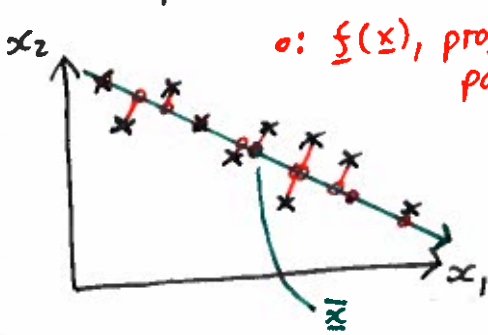
{ low-dim projected data
{ minimize

← maximize

Average over all examples

Not examinable

PCA special case $D=2, k=1$



↑
Not the principal component!

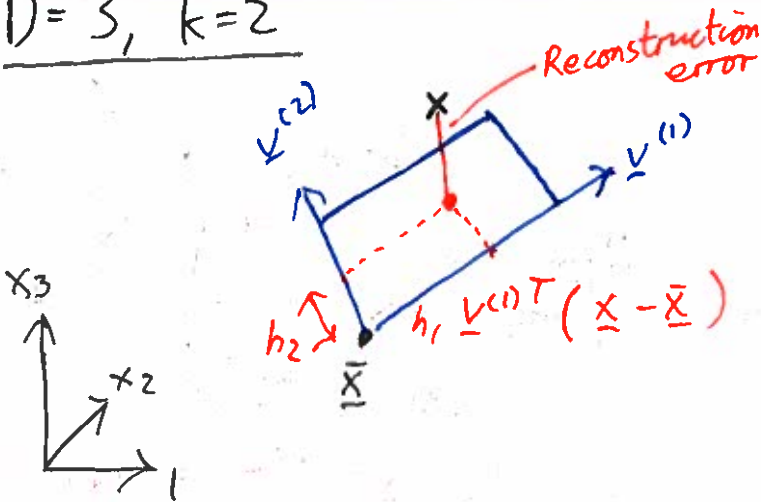
Principal component →

Minimizes $\sum_n | \cdot |^2$ — sum square distances

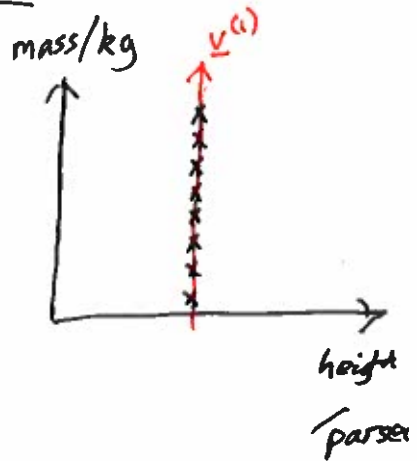
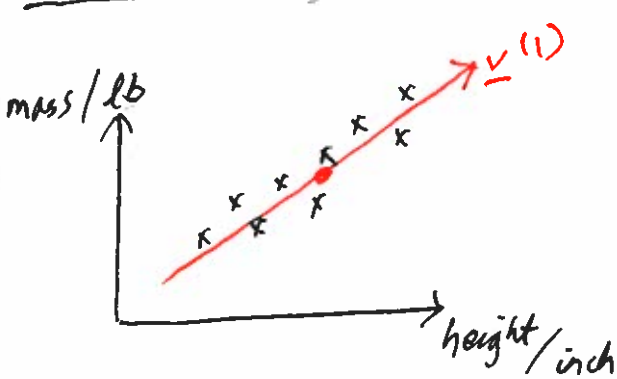
Maximizes $\text{var}(\cdot)$ — spread along line

} Don't need to prove in this course

$D=3, k=2$



Units of Features matter



Gaussian model for PCA

Assume there is process in k -dimensions

$$\underline{h}^{(n)} \sim N(\underline{0}, \mathbb{I}_k)$$

$k \times 1$

$$\underline{x}^{(n)} = V \underline{h}^{(n)} + \underline{\mu} + \text{Gaussian noise}$$

$D \times 1$ $D \times k$ $k \times 1$

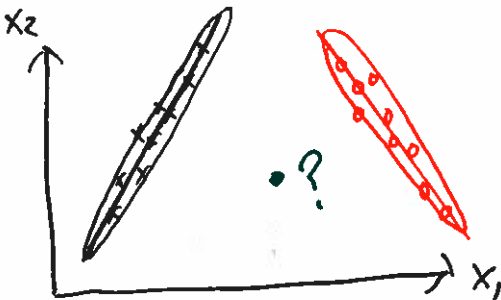
Gaussian noise
zero mean
cov. $\sigma_{\text{noise}}^2 \mathbb{I}_D$

Model \Rightarrow

$$\underline{x} \sim N(\underline{\mu}, VV^T + \sigma_{\text{noise}}^2 \mathbb{I}_D)$$

$$\begin{aligned} & \text{cov}[V\underline{h}] \\ &= \mathbb{E}[V\underline{h} \underline{h}^T V^T] \\ &= V \mathbb{E}[\underline{h} \underline{h}^T] V^T \\ &= V \mathbb{I}_k V^T \\ &= VV^T \end{aligned}$$

$D \times k$ $k \times k$ $k \times D$



x class 1
o class 0

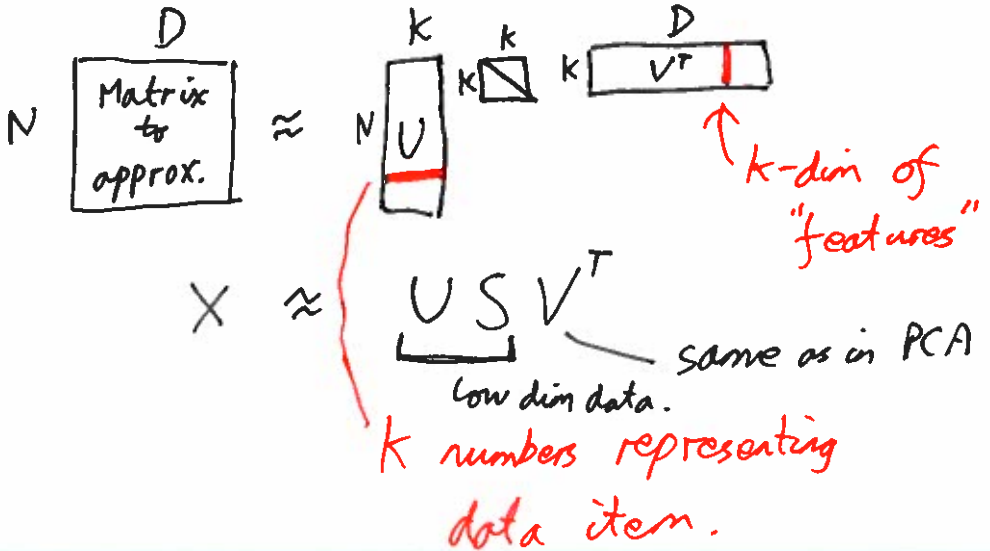
$$\begin{aligned} p(y=1 | \underline{x}=\bullet) &= \frac{0}{0+0} \\ &= \text{"NaN"} \\ &= \text{some number} \end{aligned}$$

"Probabilistic PCA"

"Factor Analysis", noise is a diagonal matrix

Truncated SVD

SVD is a standard linear algebra method



Truncated SVD

$$\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1D} \\ X_{21} & X_{22} & \cdots & X_{2D} \\ X_{31} & X_{32} & \cdots & X_{3D} \\ X_{41} & X_{42} & \cdots & X_{4D} \\ X_{51} & X_{52} & \cdots & X_{5D} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{ND} \end{bmatrix} \approx$$

```
% PCA via SVD,
% for zero-mean X:
[U, S, V] = svd(X, 0);
U = U(:, 1:K);
S = S(1:K, 1:K);
V = V(:, 1:K);
X_kdim = U*S;
X_proj = U*S*V';
```

$$\begin{bmatrix} U_{11} & \cdots & U_{1K} \\ U_{21} & \cdots & U_{2K} \\ U_{31} & \cdots & U_{3K} \\ U_{41} & \cdots & U_{4K} \\ U_{51} & \cdots & U_{5K} \\ \vdots & \ddots & \vdots \\ U_{N1} & \cdots & U_{NK} \end{bmatrix} \begin{bmatrix} S_{11} & 0 & 0 \\ 0 & \cdots & 0 \\ 0 & 0 & S_{KK} \end{bmatrix} \begin{bmatrix} V_{11} & V_{21} & \cdots & V_{D1} \\ \vdots & \vdots & \ddots & \vdots \\ V_{1K} & V_{2K} & \cdots & V_{DK} \end{bmatrix}$$

$$X \approx U \quad S \quad V^T$$

NETFLIX

Netflix Prize

Home | Rules | Leaderboard | Register | Update | Submit | Download

NETFLIX

Browse Recommendations Friends Queue Buy DVDs

Home Genres New Releases Previews Netflix Top 100 Cnt

Movies For You

Randy, the following movies were chosen based on your interest in:
[Howling for Columbine](#)
[Carnivale: Season 1](#)
[Greenheit 2003](#)

The Big One

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

★★★★★

All Discs guaranteed!

You really liked it...

Now only for just \$5.99

Show as low as

Original art

Other

Light Blue

Add

Not interested

Not interested

Red Eye

Bear Window

Give a friend

Welcome!

The Netflix Prize seeks to substantially improve the accuracy of predictions about how much someone is going to love a movie based on their movie preferences. Improve it enough and you win one (or more) Prizes. Winning the Netflix Prize improves our ability to connect people to the movies they love.

Read the [Rules](#) to see what is required to win the Prizes. If you are interested in joining the quest, you should [register a team](#).

You should also read the [frequently-asked questions](#) about the Prize. And check out how various teams are doing on the [Leaderboard](#).

Good luck and thanks for helping!

[FAQ](#) | [Forum](#) | [Netflix Home](#)

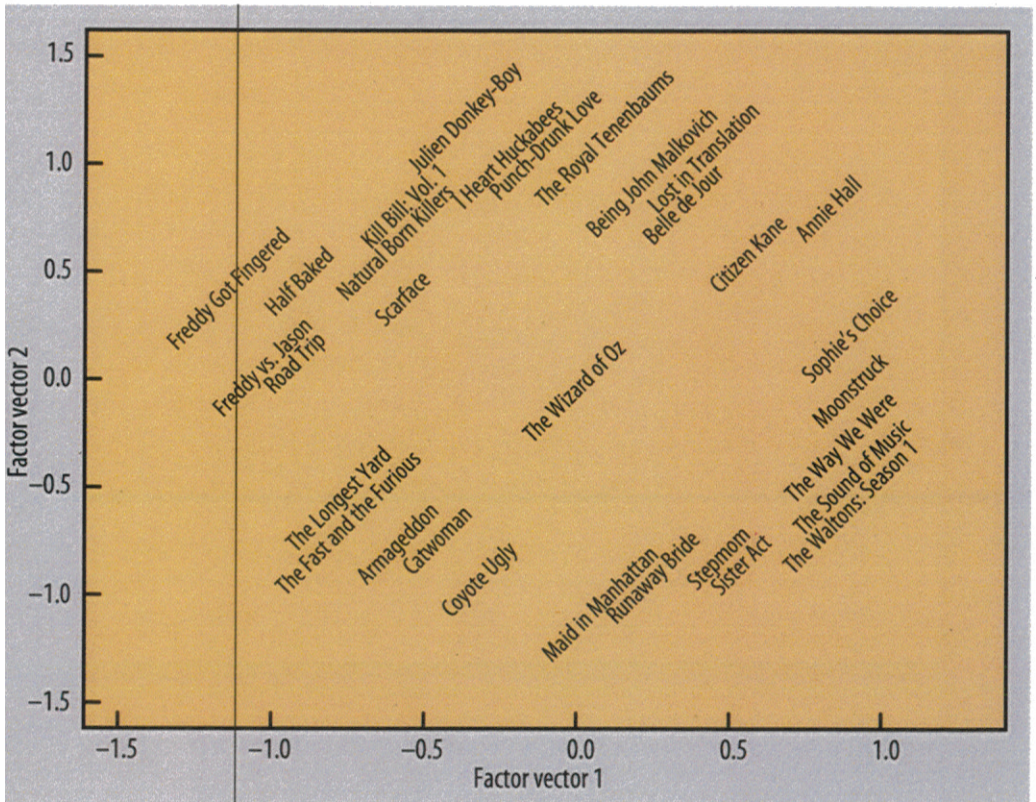
© 1997-2006 Netflix, Inc. All rights reserved.

Leaderboard

Showing Test Score. [Click here to show quiz score](#)

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries I	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace	0.8612	9.59	2009-07-24 17:18:43
9	Feeds2	0.8622	9.48	2009-07-12 13:11:51
10	BigChaos	0.8623	9.47	2009-04-07 12:33:59
11	Opera Solutions	0.8623	9.47	2009-07-24 00:34:07
12	BellKor	0.8624	9.46	2009-07-26 17:19:11
Progress Prize 2008 - RMSE = 0.8627 - Winning Team: BellKor in BigChaos				
13	xiangliang	0.8642	9.27	2009-07-15 14:53:22
14	Gravily	0.8643	9.26	2009-04-22 18:31:32
15	Ces	0.8651	9.18	2009-06-21 19:24:53
16	Invisible Ideas	0.8653	9.15	2009-07-15 15:53:04
17	Just a guy in a garage	0.8662	9.06	2009-05-24 10:02:54
18	J Dennis Su	0.8666	9.02	2009-03-07 17:16:17
19	Craig Carmichael	0.8666	9.02	2009-07-25 16:00:54
20	acmehll	0.8668	9.00	2009-03-21 16:20:50
21	MonteCarlo	0.8669	8.99	2009-03-24 10:45:14
22	IDEA2	0.8669	8.99	2009-03-25 15:37:59
23	just_a_student	0.8675	8.92	2009-07-17 08:37:11
24	Howbert	0.8677	8.90	2009-07-26 07:13:00
25	My Brain and His Chain	0.8678	8.89	2008-09-30 02:19:47
26	NewmanI	0.8681	8.86	2009-07-26 14:31:51
27	When Gravity and Dinosaurs Unite	0.8686	8.81	2008-02-29 06:48:56
28	Newman and Kramer I	0.8691	8.76	2009-07-23 23:46:20
29	Wojtek Kulik	0.8693	8.73	2009-07-24 14:20:15
30	See No Evil	0.8696	8.70	2009-07-20 13:58:43
31	netflixwinner	0.8697	8.69	2009-07-22 16:49:07
32	bbame	0.8699	8.67	2009-07-25 05:17:26
33	Peterrock	0.8702	8.64	2009-07-24 09:36:29
34	basho	0.8703	8.63	2009-07-26 00:44:48
35	bostonguy	0.8706	8.60	2009-07-24 08:52:22
36	krazy kanary	0.8708	8.58	2009-07-23 05:28:00
37	Jim Beckman	0.8709	8.57	2009-04-13 02:10:58
38	blednotik	0.8711	8.55	2009-04-18 23:28:42
39	md	0.8714	8.51	2009-01-19 17:53:42
40	GreenCircle	0.8714	8.51	2009-02-17 21:01:19

217	black_tea	0.8898	6.58	2008-11-03 20:49:06
218	lucky_13	0.8901	6.55	2008-10-09 17:09:17
219	blue_cup	0.8901	6.55	2008-10-09 23:19:05
220	DandA	0.8902	6.54	2009-01-31 14:19:12
221	OS_001	0.8905	6.51	2008-03-04 08:31:02
222	wmf2008	0.8905	6.51	2009-07-24 15:03:22
223	Bozo_The_Clown	0.8908	6.48	2007-04-06 17:50:49
224	Remco	0.8909	6.47	2007-04-02 15:53:02
225	swallow	0.8909	6.47	2008-01-25 03:26:34
226	faceupdown	0.8910	6.46	2008-11-05 03:15:56
227	Adam Bull	0.8910	6.46	2009-04-28 22:26:49
228	pawko	0.8911	6.45	2009-07-26 18:26:48
229	SteveYoo	0.8913	6.43	2009-05-31 16:01:37
230	precious_diamond	0.8914	6.41	2008-08-12 21:31:31
231	Intelligent Agent	0.8915	6.40	2009-06-03 16:22:30
232	niko	0.8917	6.38	2008-12-13 17:21:56
233	MaximisedExpectations	0.8918	6.37	2009-07-05 13:25:01
234	sweaterr	0.8919	6.36	2009-04-02 18:39:13
235	Asylum Residents	0.8920	6.35	2007-06-14 17:45:14
236	vdicarlo	0.8921	6.34	2007-03-25 05:12:17
237	simonfunk	0.8921	6.34	2007-03-27 04:10:45
238	top_112	0.8921	6.34	2008-01-23 02:56:14
239	Startibartfast	0.8922	6.33	2008-02-20 05:09:34
240	nemo	0.8923	6.32	2008-01-22 03:20:32
241	richard	0.8924	6.31	2008-01-23 02:47:07
242	PeanutButterLovers	0.8924	6.31	2009-07-07 18:23:13
243	Magnificent7	0.8926	6.29	2009-07-26 14:25:03
244	ahan	0.8927	6.28	2007-09-27 20:05:59
245	mokidao	0.8927	6.28	2007-09-27 20:19:57
246	patience	0.8928	6.27	2008-09-18 03:59:43
247	Team2403	0.8930	6.25	2008-03-12 07:07:10
248	green_tea	0.8931	6.24	2008-11-03 20:17:20
249	kouburgs	0.8932	6.23	2009-06-29 12:48:47
250	top ranker	0.8934	6.20	2008-02-29 18:12:48
251	Jaime_Carsten	0.8934	6.20	2008-09-09 21:12:00
252	CS_JH	0.8935	6.19	2008-09-09 20:36:44
253	Schumarzi	0.8937	6.17	2008-02-22 01:59:04
254	Christinas_Team	0.8939	6.15	2008-09-09 20:13:08
255	Pattern Excavator	0.8940	6.14	2007-11-02 16:22:01
256	Smultron	0.8944	6.10	2009-02-09 20:13:08
257	Need a job	0.8945	6.09	2009-07-26 05:48:01
258	Gerald Schwab	0.8946	6.08	2009-07-05 18:39:03
259	Zazpiak Bat	0.8948	6.06	2007-06-14 21:04:51
260	Learnflix	0.8948	6.06	2007-07-02 21:17:37
261	Atomic Rapadura	0.8950	6.04	2007-06-10 21:54:29
262	pgolle	0.8950	6.04	2007-09-11 05:08:49
263	Donkey Power	0.8950	6.04	2007-09-13 05:31:00
264	NYU Michael	0.8950	6.04	2009-07-25 14:04:58
265	Witi	0.8952	6.02	2007-09-09 14:28:53
266	The Gaussian Screen	0.8952	6.02	2007-10-23 21:25:16
267	W	0.8952	6.02	2007-11-28 17:36:34
268	Lucky VT	0.8952	6.02	2009-07-24 13:05:19
269	Rapadura	0.8956	5.97	2007-07-06 15:03:48
270	beluga	0.8956	5.97	2007-12-26 06:36:20
271	Rego Park	0.8957	5.96	2009-07-20 15:09:15
272	Dori.	0.8958	5.95	2009-06-28 04:33:34
273	Doron Rippel	0.8958	5.95	2009-07-03 21:41:42
274	DGrant	0.8958	5.95	2009-07-03 21:50:59
275	Spring Force	0.8959	5.94	2007-05-04 16:29:40



From Koren et al. (2009)