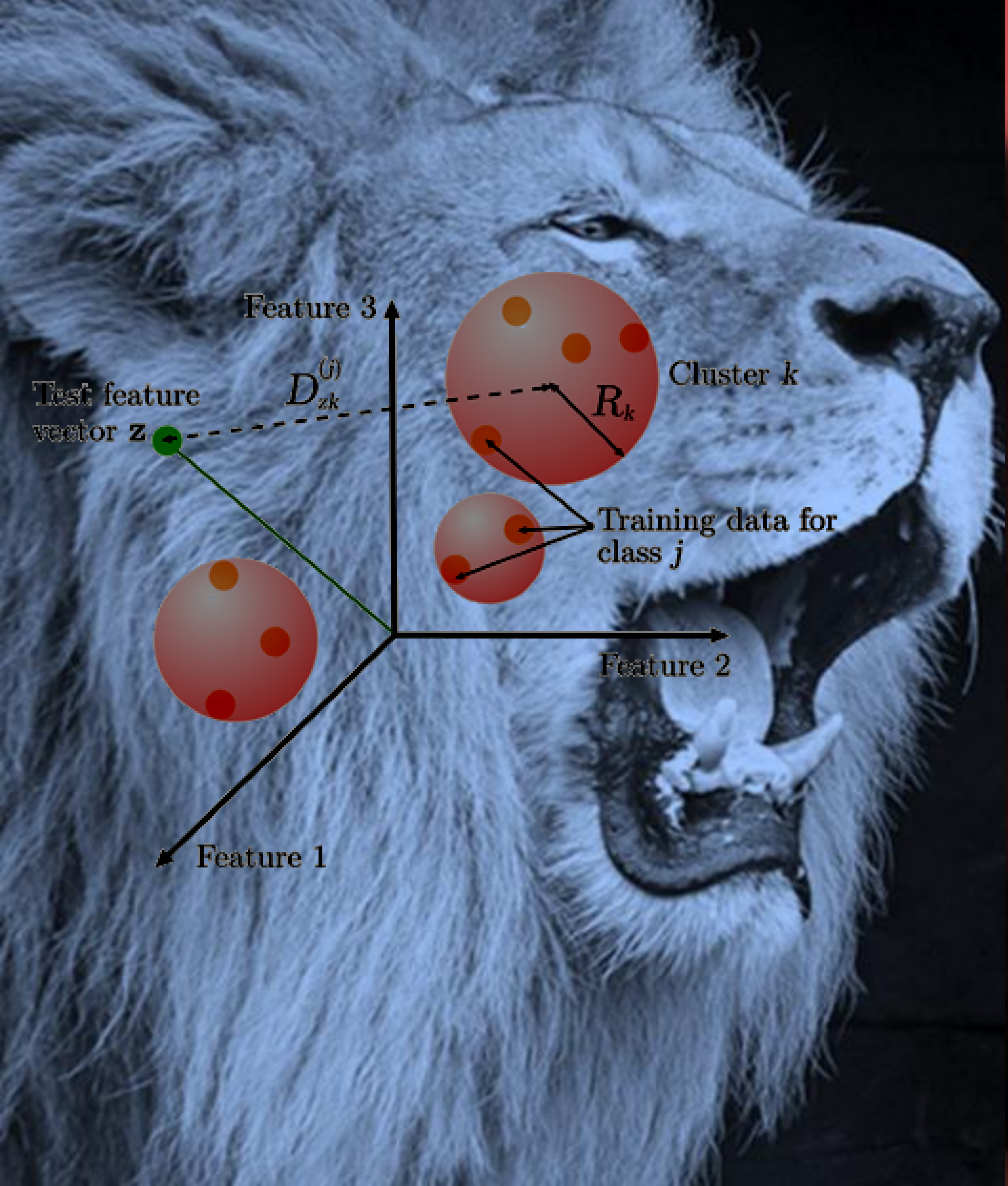
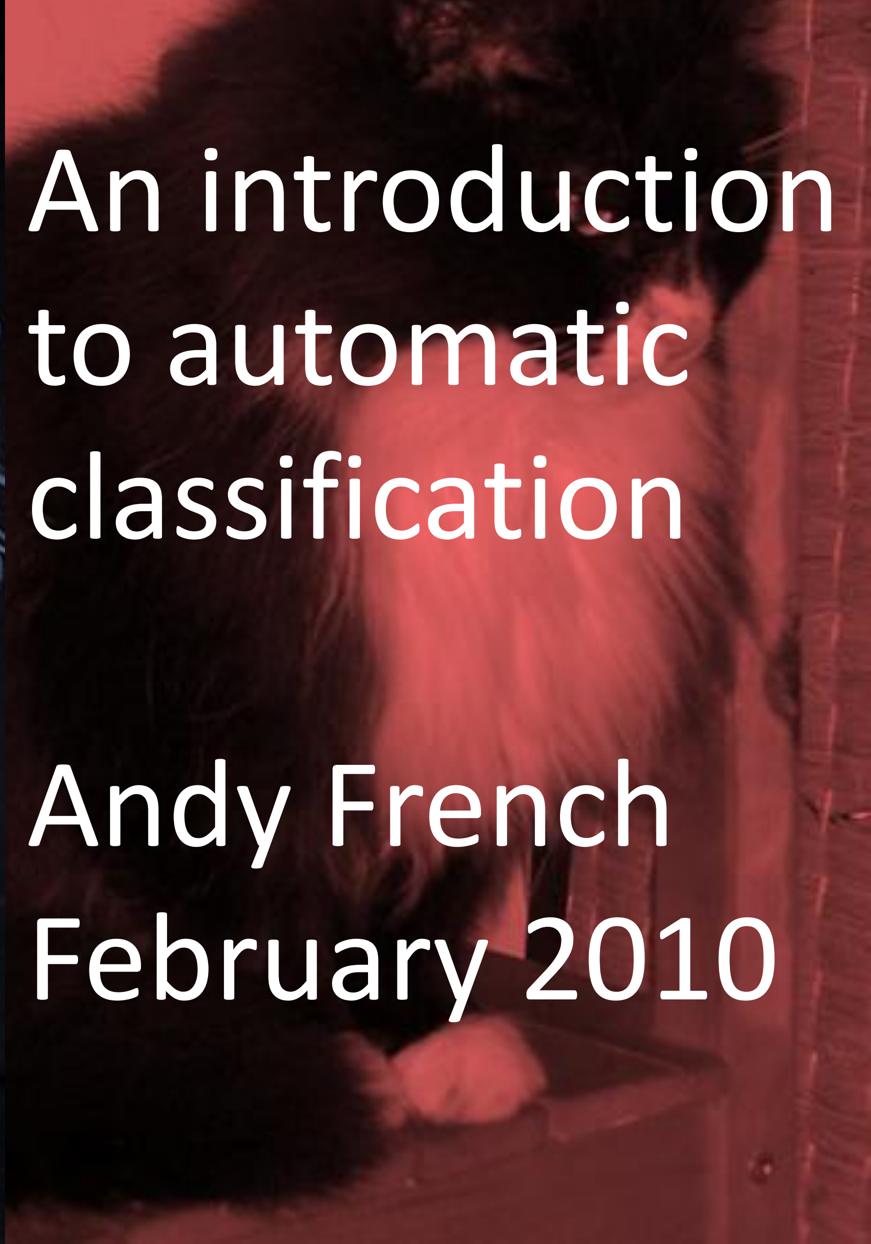
$$g_i(\mathbf{x}) = -\frac{1}{2}\left(\mathbf{x}-\mathbf{m}_i\right)^T \mathbf{\Sigma}_i^{-1}\left(\mathbf{x}-\mathbf{m}_i\right) - \frac{1}{2}\log\left(|\mathbf{\Sigma}_i|\right) + \log\left(\frac{Z_i}{Z}\right)$$





An introduction to automatic classification

Andy French
February 2010

Target recognition “at a glance”

“One of the most potent of human skills is the ability to rapidly recognize and ***classify*** *environmental stimuli, often when such signals are severely corrupted. Of this toolkit of sensors and processing, the method of visual facial recognition is perhaps the most impressive. Typically, a successful recognition (i.e. a name attached) will occur in* ***120 ms****, with cruder classifications (for example classification of a species group from a background) in as little as* ***50ms****.”*

Can a machine be built with this level of performance?

This ‘introduction’ is but a glance at a large and active research area!

For a *much* more complete introduction see

Duda, R.O., Hart, P.E, Stork, D.G., Pattern Classification. 2nd Edition. John Wiley & Sons Inc. 2001.

Webb. A., Statistical Pattern Recognition. 2nd Edition. John Wiley & Sons Ltd. 2002.

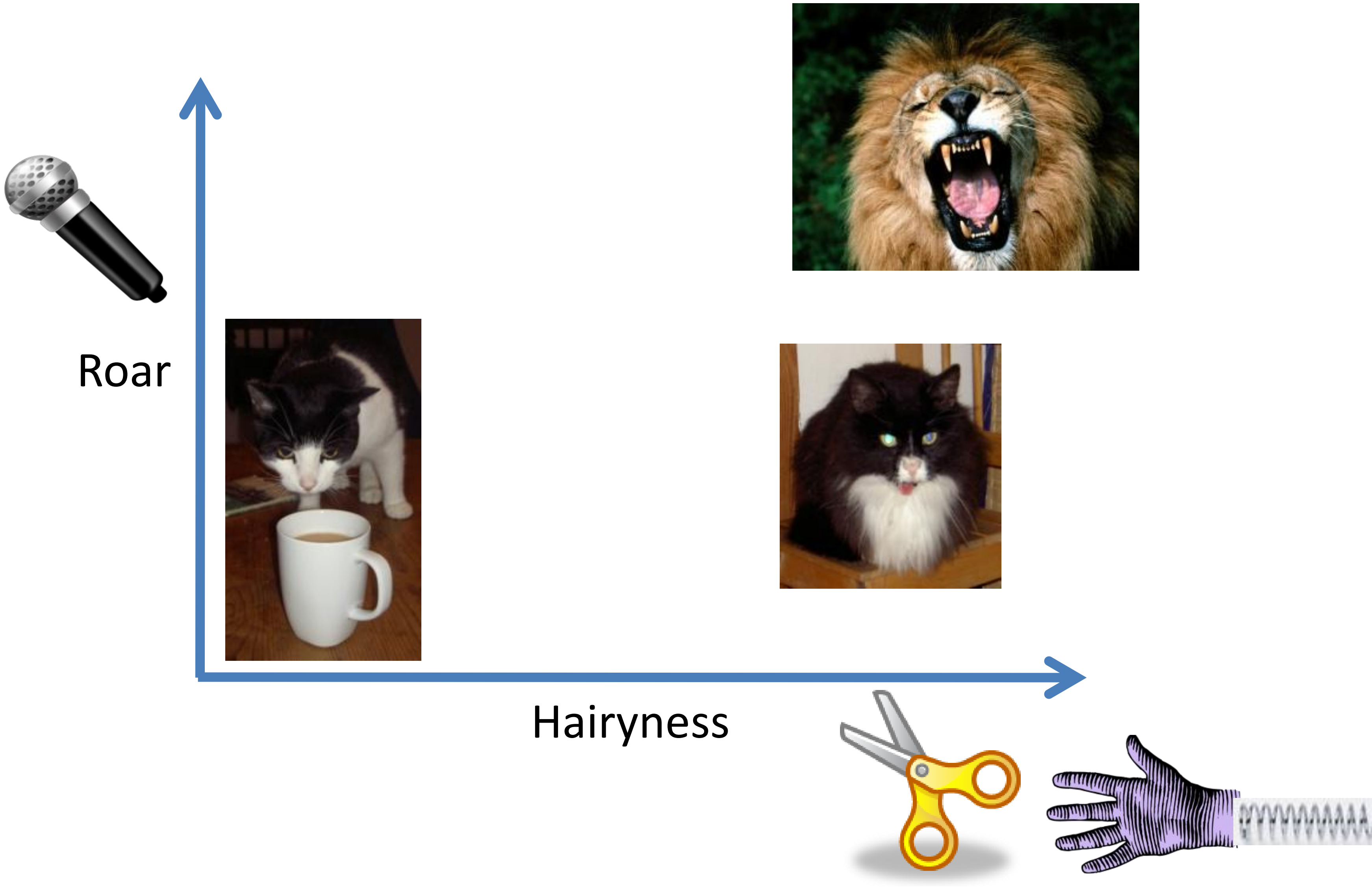
Let us start in a similar fashion to Duda, with a practical example of a classification problem…

A problem of cat classification….

There are many cats out there

How can I be sure to let the right one in…?

Mechanized Entrance Test Of Cats



Feature measurements

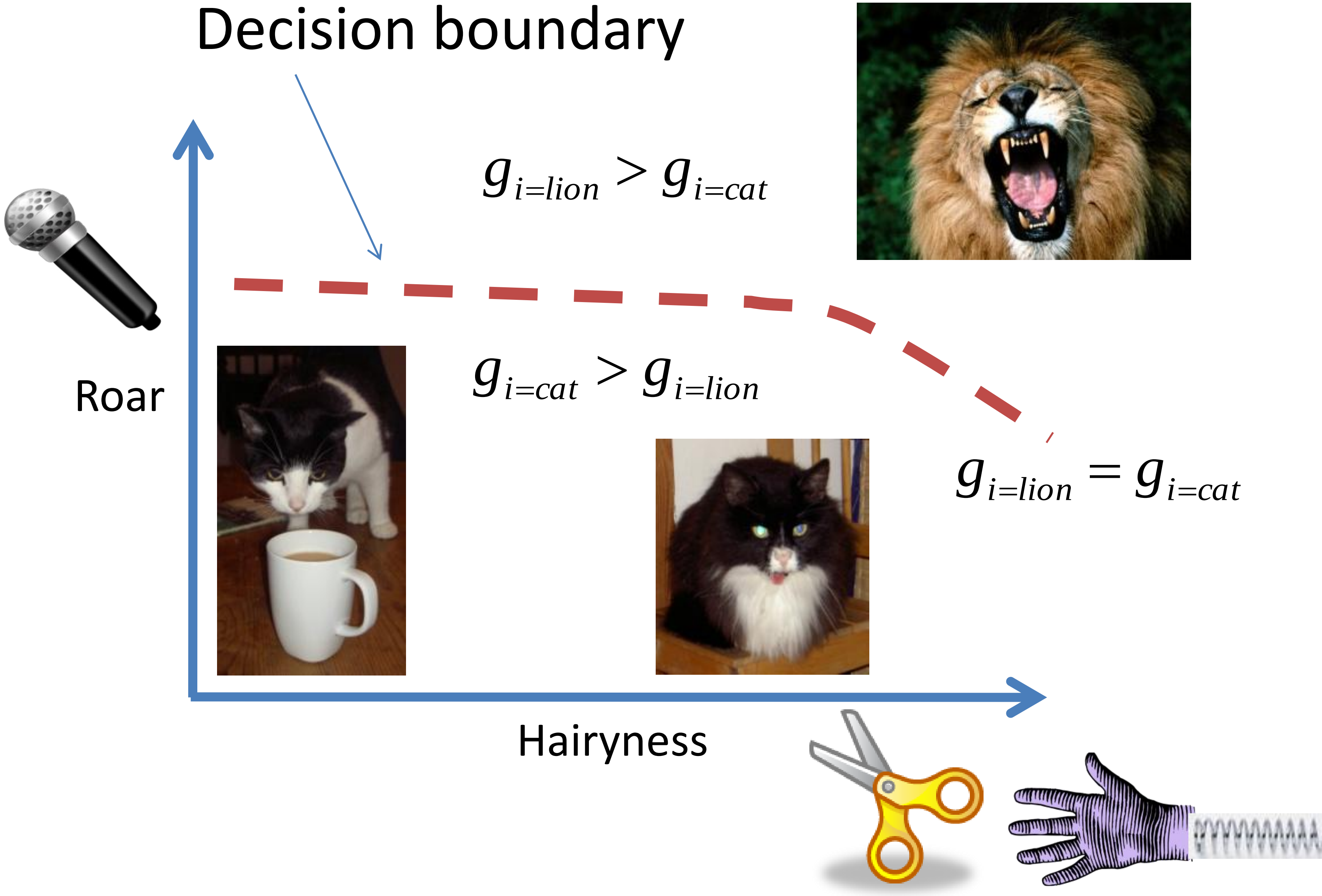
✂ 🎤

$$g_{i=cat} = f_{cat}(\text{hairy ness,roar})$$

$$g_{i=lion} = h_{lion}(\text{hairy ness,roar})$$

Class label

Classifier *discriminant* functions



Decision boundary

$g_{i=lion} > g_{i=cat}$

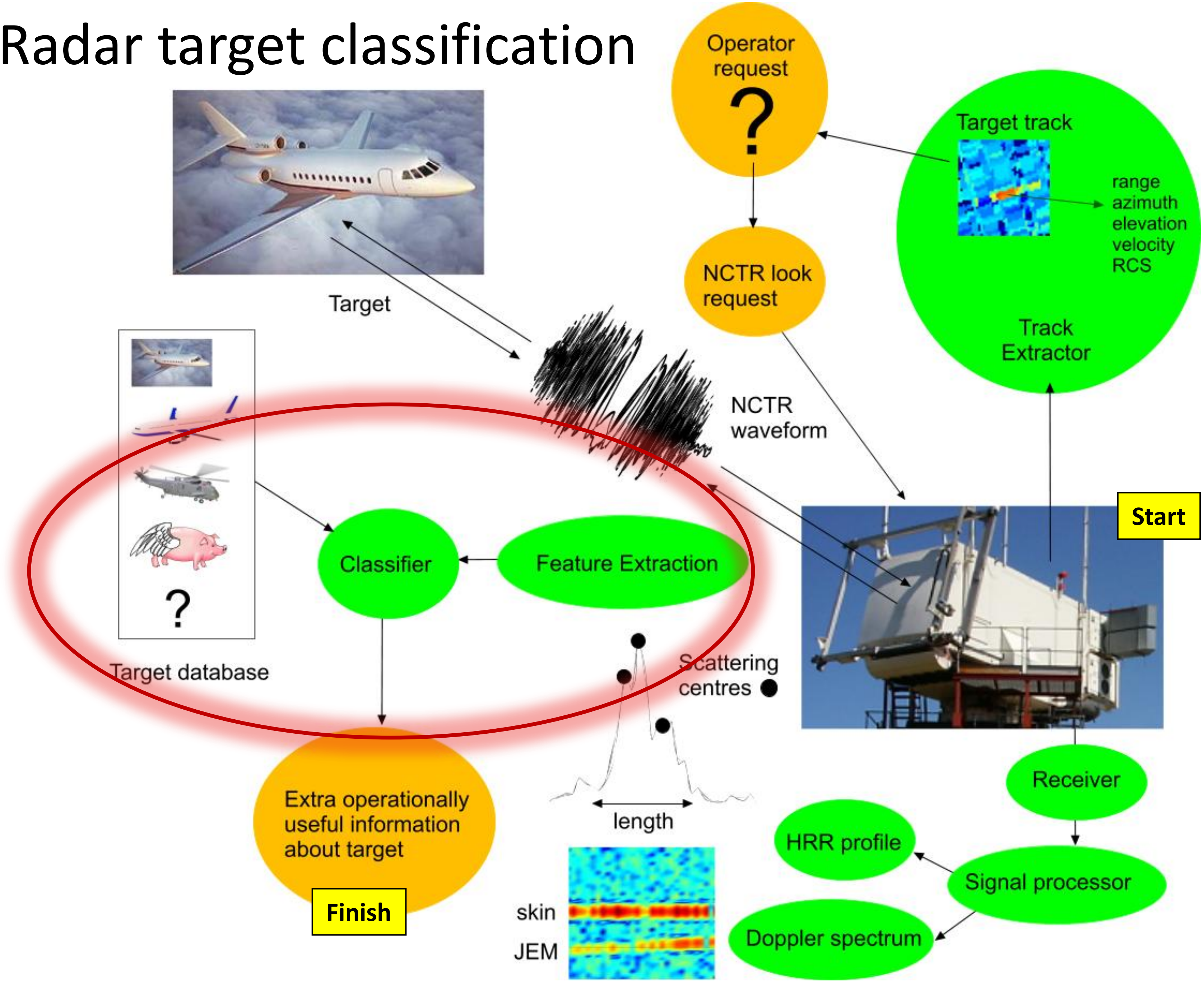
Roar

$g_{i=cat} > g_{i=lion}$

$g_{i=lion} = g_{i=cat}$

Hairyness

Radar target classification



Operator request

?

Target track

range
azimuth
elevation
velocity
RCS

Track Extractor

NCTR look request

Target

NCTR waveform

Start

Target database

?

Classifier

Feature Extraction

Scattering centres ●

length

Receiver

Signal processor

HRR profile

Doppler spectrum

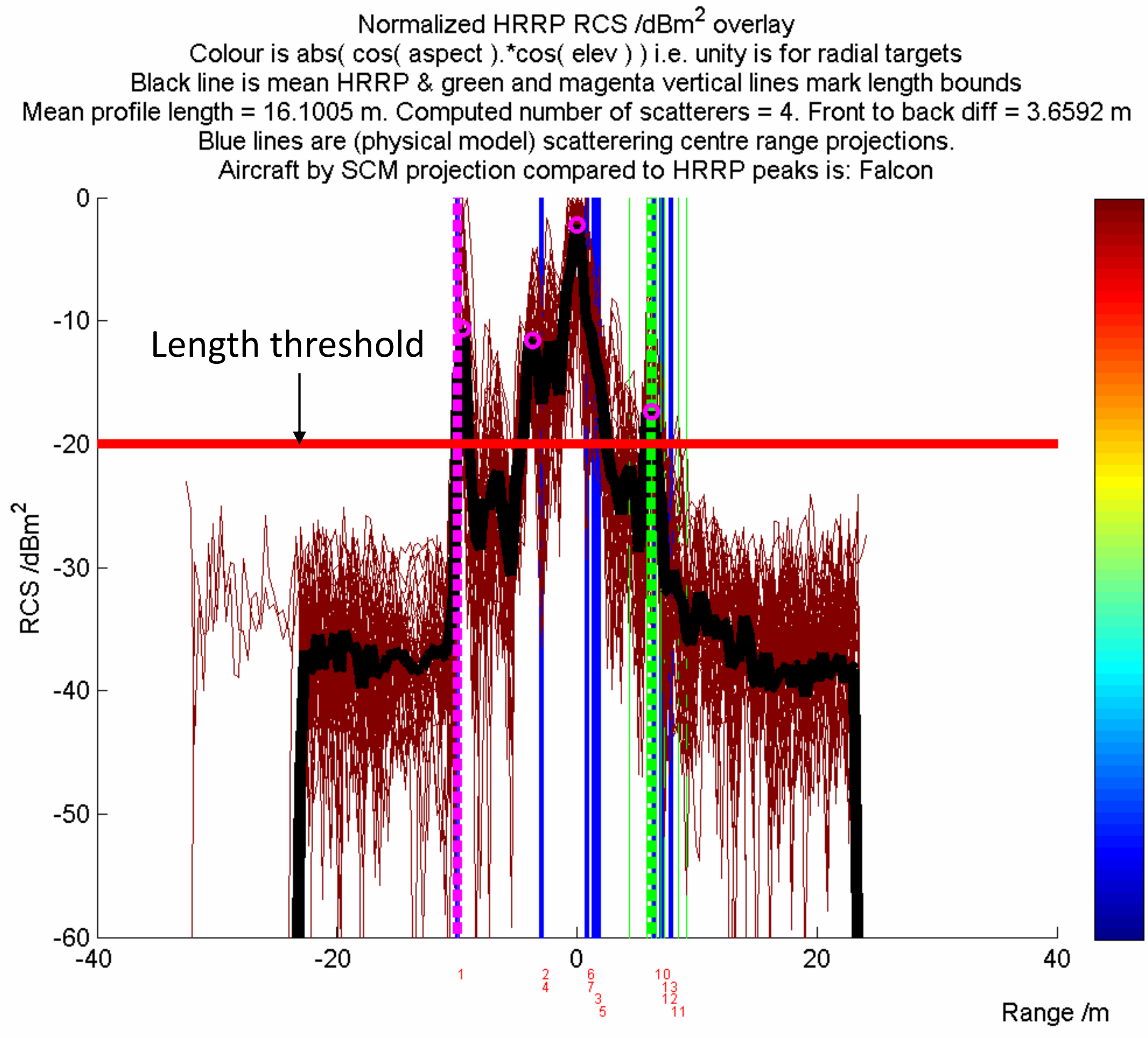
skin

JEM

Extra operationally useful information about target

Finish

Feature extraction: Radar length

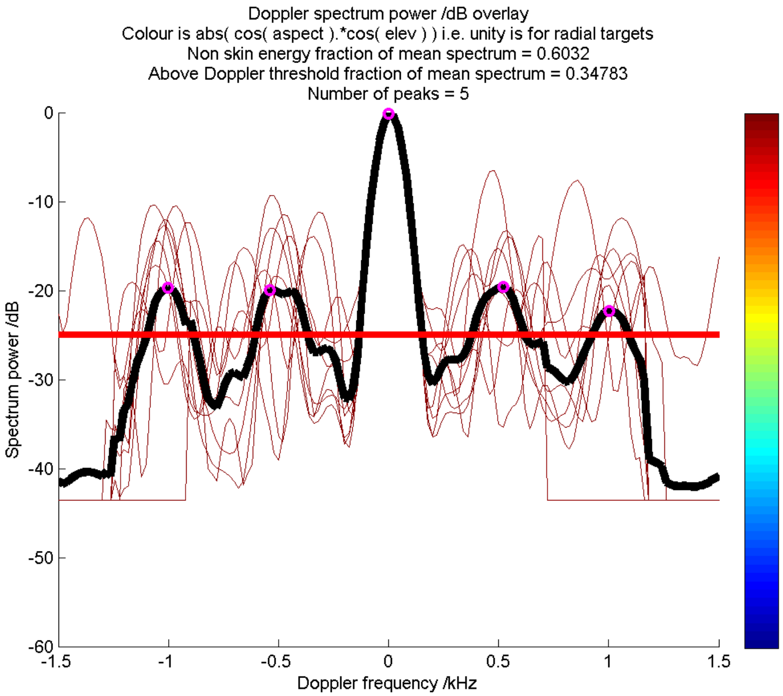


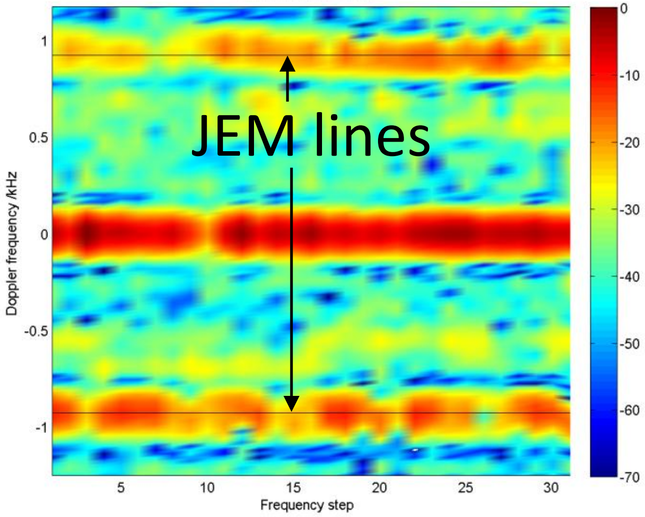
Normalized HRRP RCS /dBm^2 overlay
Colour is abs(cos(aspect).*cos(elev)) i.e. unity is for radial targets
Black line is mean HRRP & green and magenta vertical lines mark length bounds
Mean profile length = 16.1005 m. Computed number of scatterers = 4. Front to back diff = 3.6592 m
Blue lines are (physical model) scatterering centre range projections.
Aircraft by SCM projection compared to HRRP peaks is: Falcon

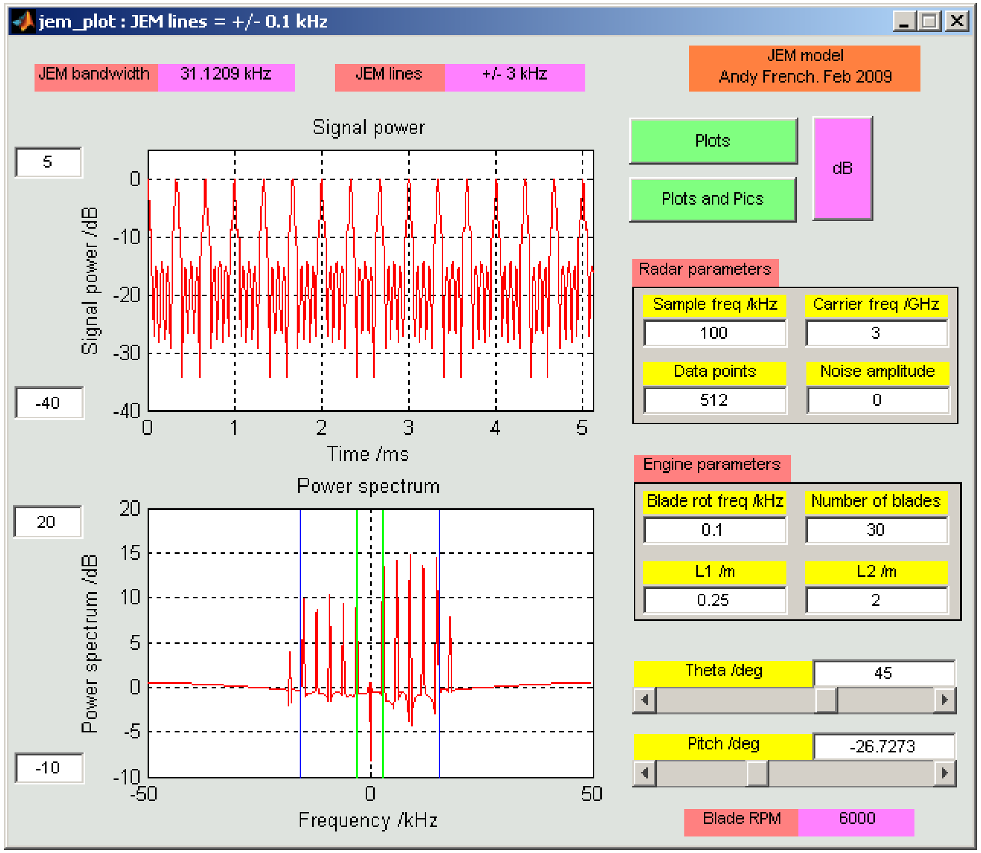
Length threshold

Inbound Falcon jet aircraft

Doppler processing: Jet Engine Modulation (JEM)






jem_plot : JEM lines = +/- 0.1 kHz

JEM bandwidth 31.1209 kHz

JEM lines +/- 3 kHz

JEM model
Andy French. Feb 2009

Radar parameters

Sample freq /kHz	Carrier freq /GHz
100	3
Data points	Noise amplitude
512	0

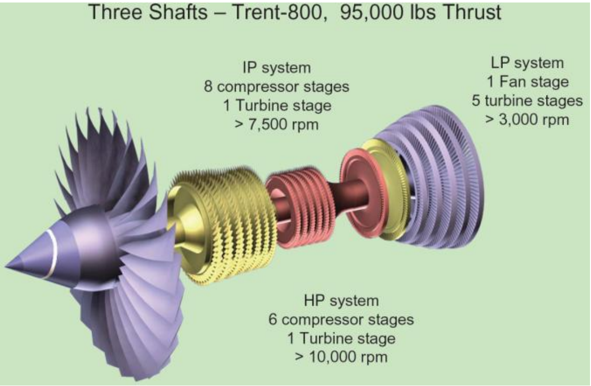
Engine parameters

Blade rot freq /kHz	Number of blades
0.1	30
L1 /m	L2 /m
0.25	2

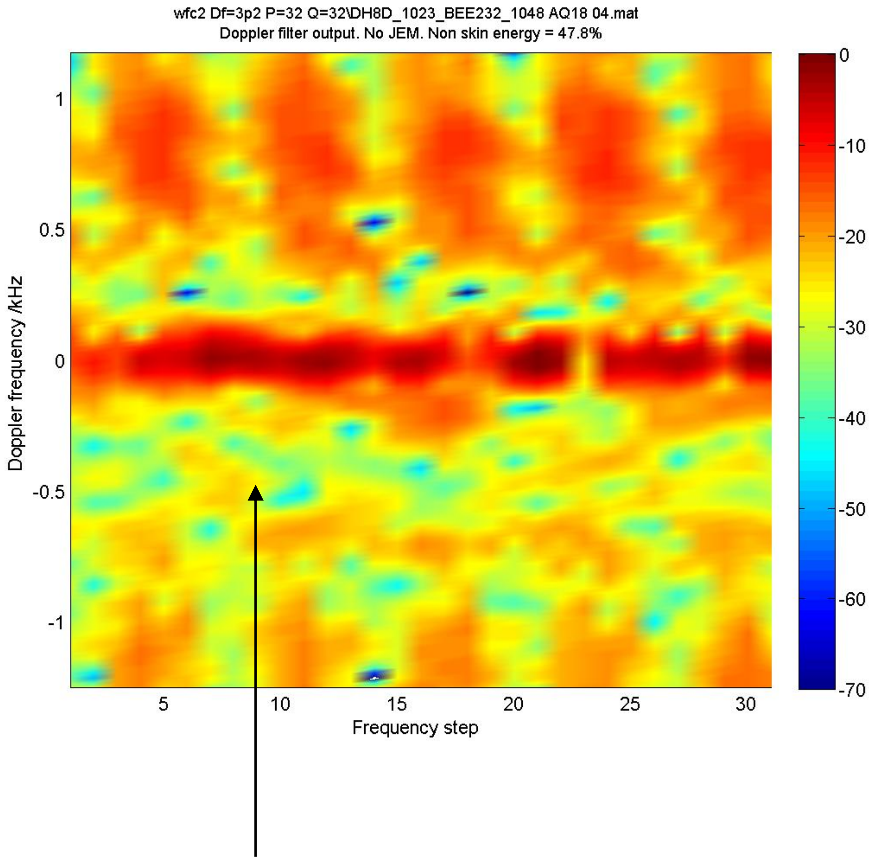
Theta /deg 45

Pitch /deg -26.7273

Blade RPM 6000



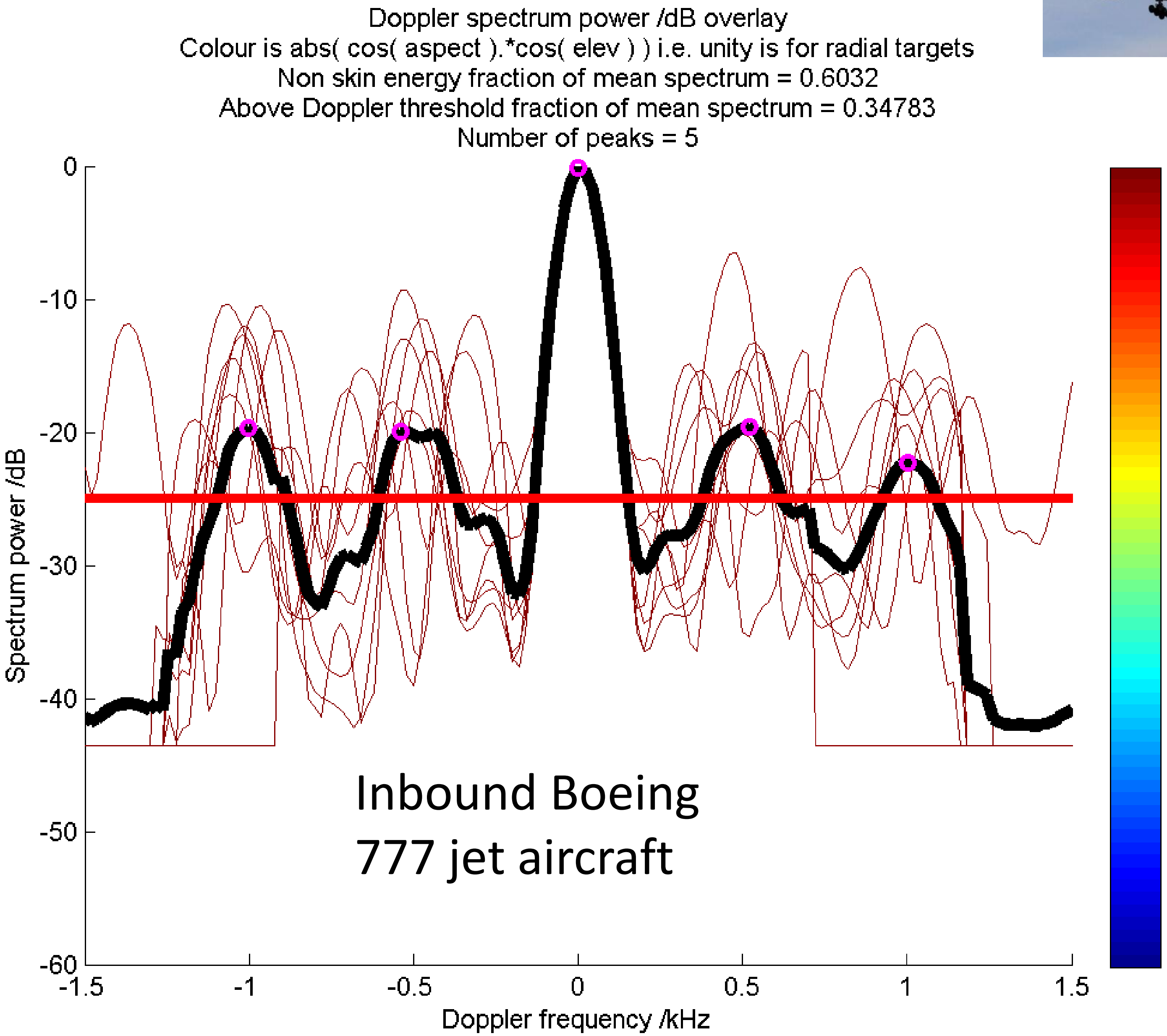
(Aside) Doppler processing: Propeller modulation



Doppler spectrum for 32 pulse, 32 frequency step waveform E 2.5kHz PRF

Dash8 six blade propeller aircraft

Feature extraction: Doppler spectra



Doppler spectrum power /dB overlay
Colour is abs(cos(aspect).*cos(elev)) i.e. unity is for radial targets
Non skin energy fraction of mean spectrum = 0.6032
Above Doppler threshold fraction of mean spectrum = 0.34783
Number of peaks = 5
Spectrum power /dB
0
-10
-20
-30
-40
-50
-60
Doppler frequency /kHz
-1.5
-1
-0.5
0
0.5
1
1.5

Inbound Boeing 777 jet aircraft

Aim: Design a classifier based upon **measured feature statistics**

Example #1: a **parametric** (**Gaussian**) classifier

i.e. feature data is assumed to adopt a *Gaussian distribution*, characterized by *mean* and *covariance* parameters

Gaussian distribution of feature vectors **x**, *given class w*$_i$

$$p(\mathbf{x} \mid w_i) = \frac{1}{(2\pi)^{\frac{M}{2}} \left|\boldsymbol{\Xi}_i\right|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}_i\right)^T \boldsymbol{\Xi}_i^{-1}\left(\mathbf{x}-\boldsymbol{\mu}_i\right)\right\}$$

mean

covariance

Apply *Bayes Theorem* to determine the *posterior probability*, which will be proportional to our desired discriminant function $g_i(\mathbf{x})$

$$\underbrace{p(w_i \mid \mathbf{x})}_{\text{posterior}} p(\mathbf{x}) = \underbrace{p(\mathbf{x} \mid w_i)}_{\text{likelihood}} \; \underbrace{p(w_i)}_{\text{prior}}$$

$$\downarrow$$

$$\boxed{g_i(\mathbf{x}) = \log\left(p(w_i)\right) - \frac{1}{2}\log\left(\left|\boldsymbol{\Xi}_i\right|\right) - \frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}_i\right)^T \boldsymbol{\Xi}_i^{-1}\left(\mathbf{x}-\boldsymbol{\mu}_i\right)}$$

Voila! The Gaussian classifier.
But how do we compute the mean and covariance from **training data**?

Given a training data set $\left\{\mathbf{t}_1^{(i)}....\mathbf{t}_{Z_i}^{(i)}\right\}$, Webb [110] pp35 shows that maximizing the *Likelihood function*

$$L\left(\left\{\mathbf{t}_1^{(i)}....\mathbf{t}_{Z_i}^{(i)}\right\} \mid \boldsymbol{\mu}_i, \Xi_i\right) = \prod_{z=1}^{Z_i} p(\mathbf{t}_z^{(i)} \mid w) = \prod_{z=1}^{Z_i} \frac{1}{(2\pi)^{\frac{M}{2}} \left|\Xi_i\right|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\left(\mathbf{t}_z^{(i)} - \boldsymbol{\mu}_i\right) \Xi_i^{-1} \left(\mathbf{t}_z^{(i)} - \boldsymbol{\mu}_i\right)\right\}$$

M is the number of features

$$\frac{\partial \log(L)}{\partial \boldsymbol{\mu}_i} = 0 \text{ and } \frac{\partial \log(L)}{\partial \boldsymbol{\Xi}_i} = 0$$

$$\boldsymbol{\mu}_i \approx \mathbf{m}_i = \frac{1}{Z_i} \sum_{z=1}^{Z_i} \mathbf{t}_z^{(i)}$$

Sample mean

Sample covariance

$$\boldsymbol{\Xi}_i \approx \boldsymbol{\Sigma}_i = \frac{1}{Z_i} \sum_{z=1}^{N} \left(\mathbf{t}_z - \mathbf{m}_i\right) \left(\mathbf{t}_z - \mathbf{m}_i\right)^T$$

Using the assignment of the prior probability to be $p(w_i) = \frac{Z_i}{Z}$ where $Z = \sum_i Z_i$, one arrives at the *Gaussian classifier*

$$\boxed{g_i(\mathbf{x}) = -\frac{1}{2}\left(\mathbf{x} - \mathbf{m}_i\right)^T \boldsymbol{\Sigma}_i^{-1} \left(\mathbf{x} - \mathbf{m}_i\right) - \frac{1}{2} \log\left(\left|\boldsymbol{\Sigma}_i\right|\right) + \log\left(\frac{Z_i}{Z}\right)}$$

Bayesian FRD classifier

(Bayesian) Friedman regularized discriminant function

$$g_i(\mathbf{x}) = -\tfrac{1}{2}(\mathbf{x}-\mathbf{m}_i)^T\left[\boldsymbol{\Sigma}_i^{\lambda,\gamma}\right]^{-1}(\mathbf{x}-\mathbf{m}_i) - \tfrac{1}{2}\log\left(\left|\boldsymbol{\Sigma}_i^{\lambda,\gamma}\right|\right) + \log\left(\frac{Z_i}{Z}\right)$$

$$\boldsymbol{\Sigma}_i^{\lambda,\gamma} = (1-\gamma)\boldsymbol{\Sigma}_i^{\lambda} + \gamma c_i(\lambda)\mathbf{I}_Z$$

$$c_i(\lambda) = \frac{\mathrm{Tr}\left(\boldsymbol{\Sigma}_i^{\lambda}\right)}{Z}$$

$$\boldsymbol{\Sigma}_i^{\lambda} = \frac{(1-\lambda)\mathbf{S}_i + \lambda\mathbf{S}}{(1-\lambda)Z_i + \lambda Z}$$

$$\mathbf{S}_i = Z_i\boldsymbol{\Sigma}_i$$

$$\mathbf{S} = Z\mathbf{S}_W$$

$$Z = \boldsymbol{\Sigma}_i Z_i$$

Computed for TRAINING vectors **t**

↓

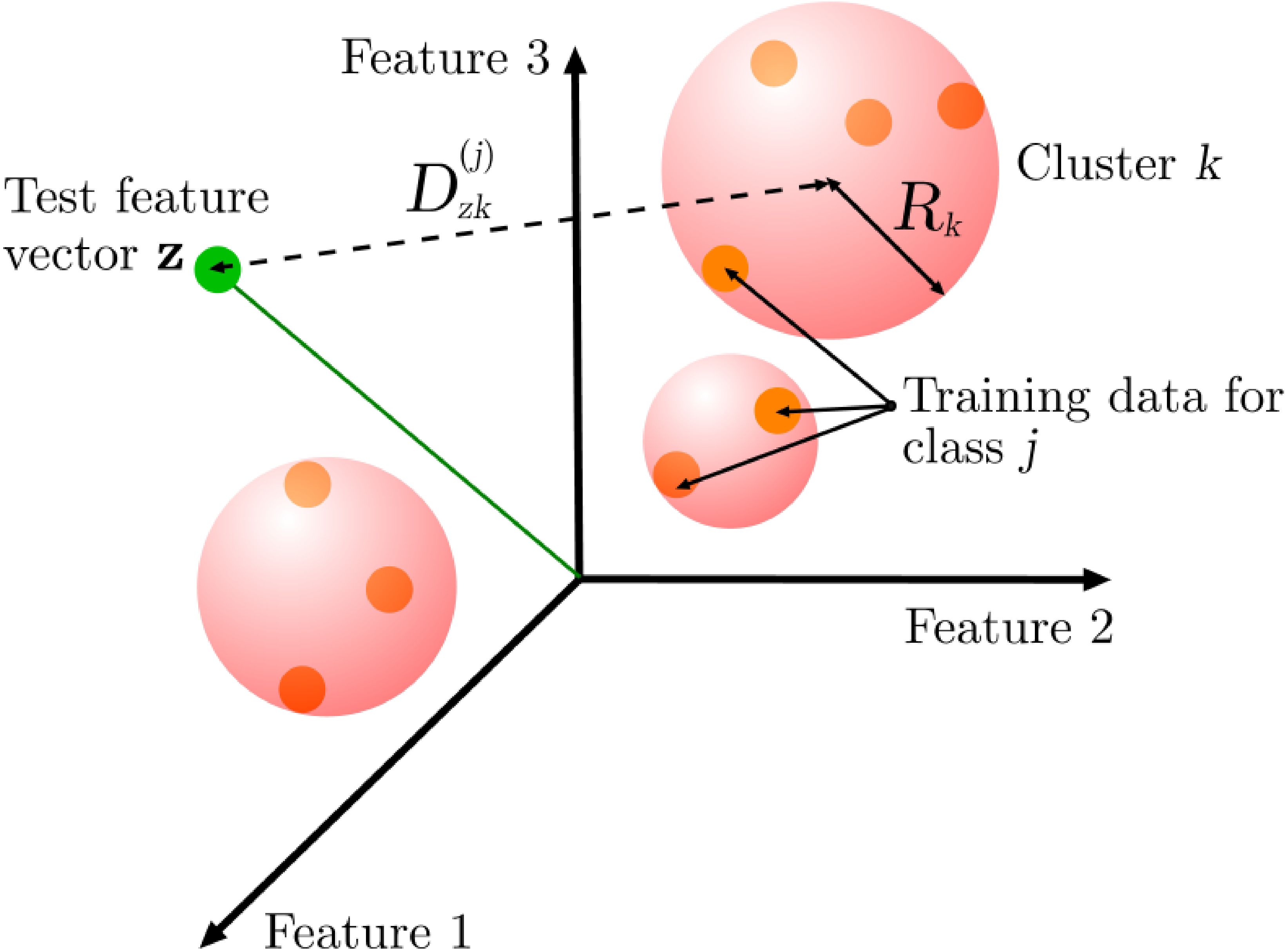
Sample within class covariance

$$\mathbf{S}_W = \sum_{i=1}^{C}\frac{Z_i}{Z}\boldsymbol{\Sigma}_i$$

$$\mathbf{S}_B = \sum_{i=1}^{C}\frac{Z_i}{Z}(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T$$

Sample between class covariance

Example #2: **k-means** *non-parametric* classifier



k-means classification does not assume an *a-priori* feature distribution. Instead one uses the **k-means clustering algorithm** to automatically group training data into *K* clusters

$$\mathbf{c}_k^{(i)} = \frac{1}{\eta_k}\sum_{z=1}^{Z_i} U_{kz}\mathbf{t}_z^{(i)}$$

Centre of cluster hyperspheres

(binary) cluster membership matrix ***U***. Start with random assignments!

Training data for class *i*

$$\eta_k = \sum_{z=1}^{Z_i} U_{kz}$$

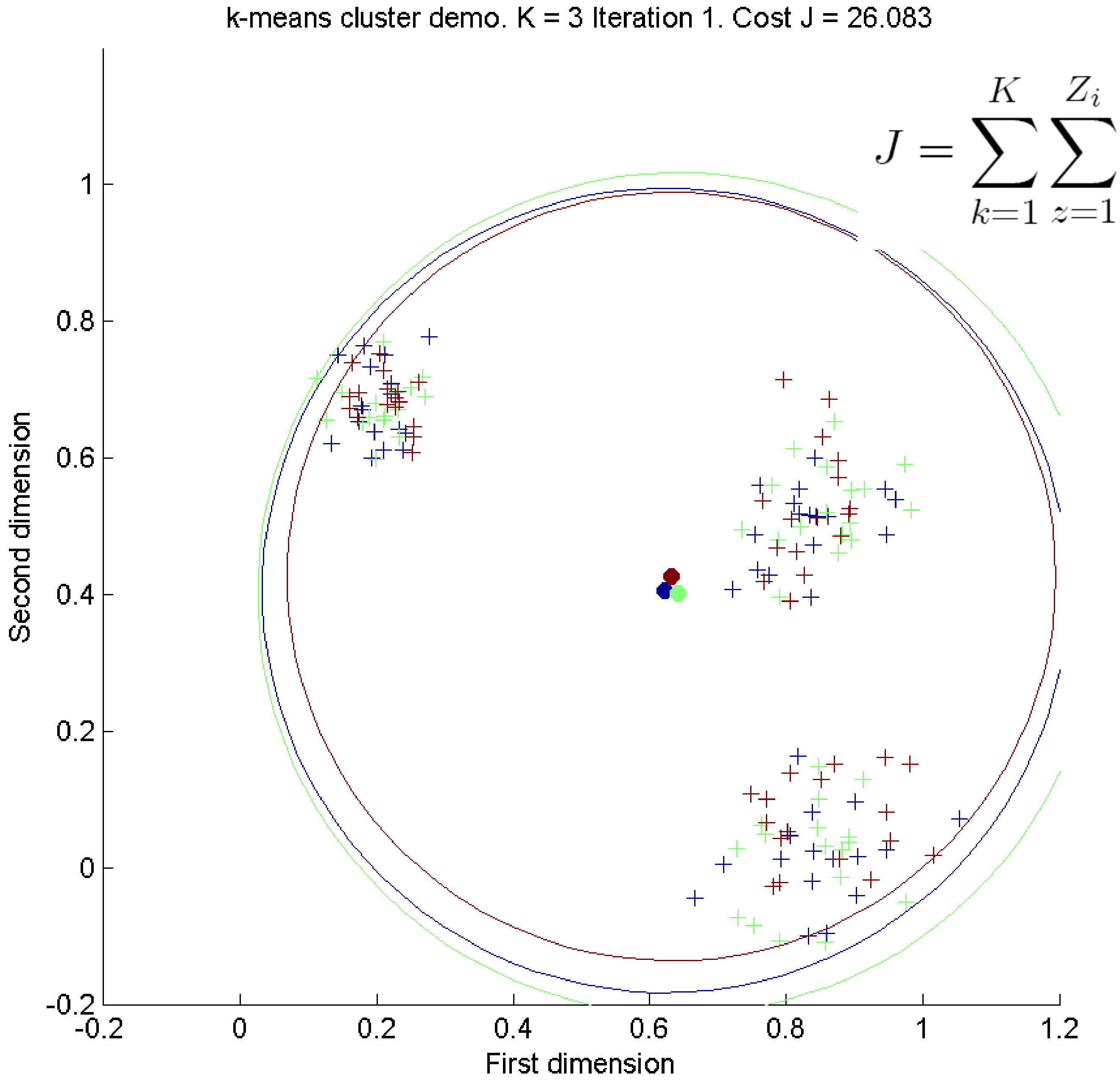
Distance between training data and cluster centres

$$[\mathbf{D}]_{kz} \equiv D_{kz} = \left|\mathbf{t}_z^{(i)} - \mathbf{c}_k^{(i)}\right|$$

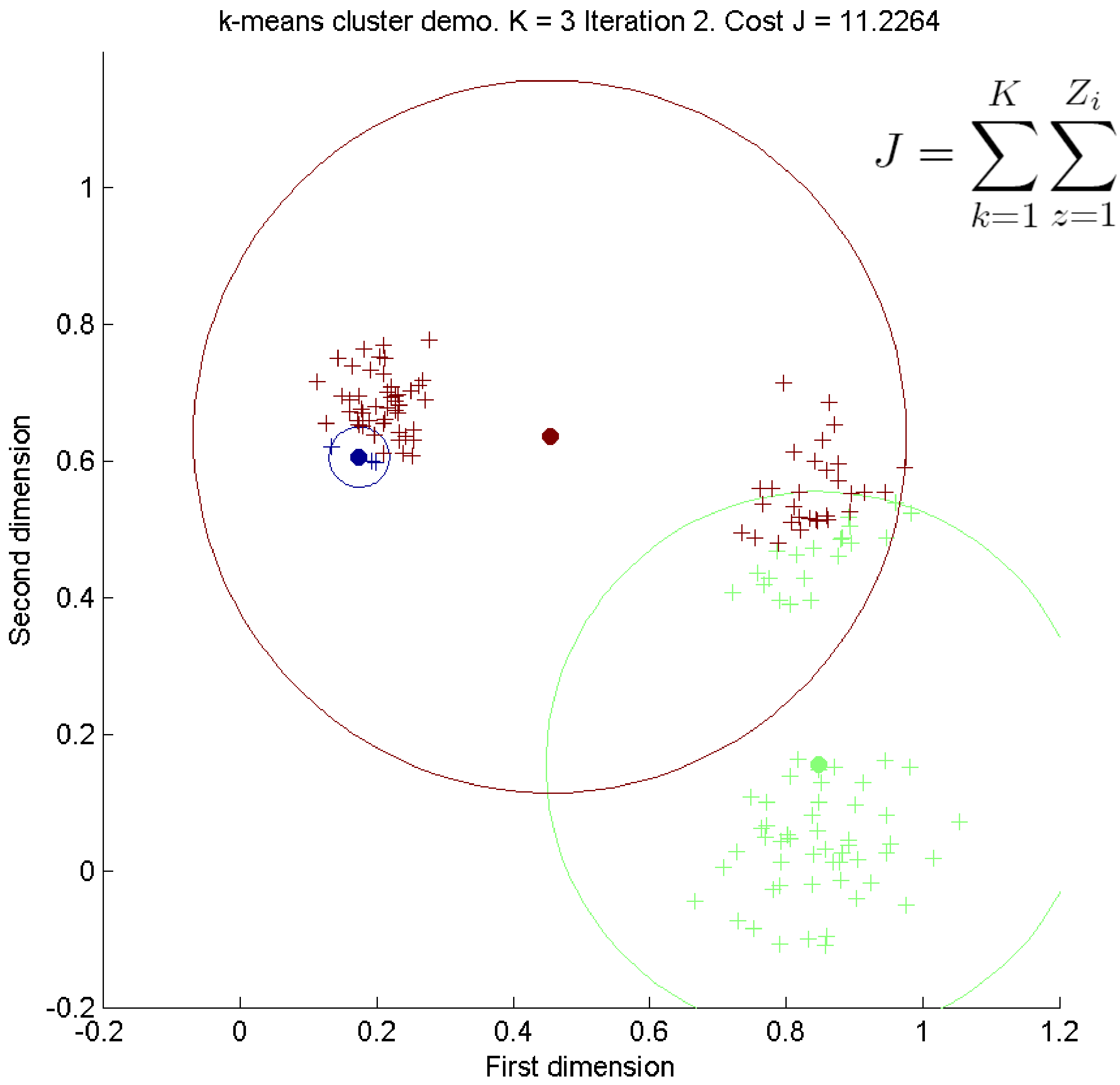
$$R_k^{(i)} = \max\left[D_{k1}U_{k1}....D_{kZ_i}U_{kZ_i}\right]$$

Radii of cluster hyperspheres

Update membership ***U*** based on nearest hypersphere centre for each training feature vector

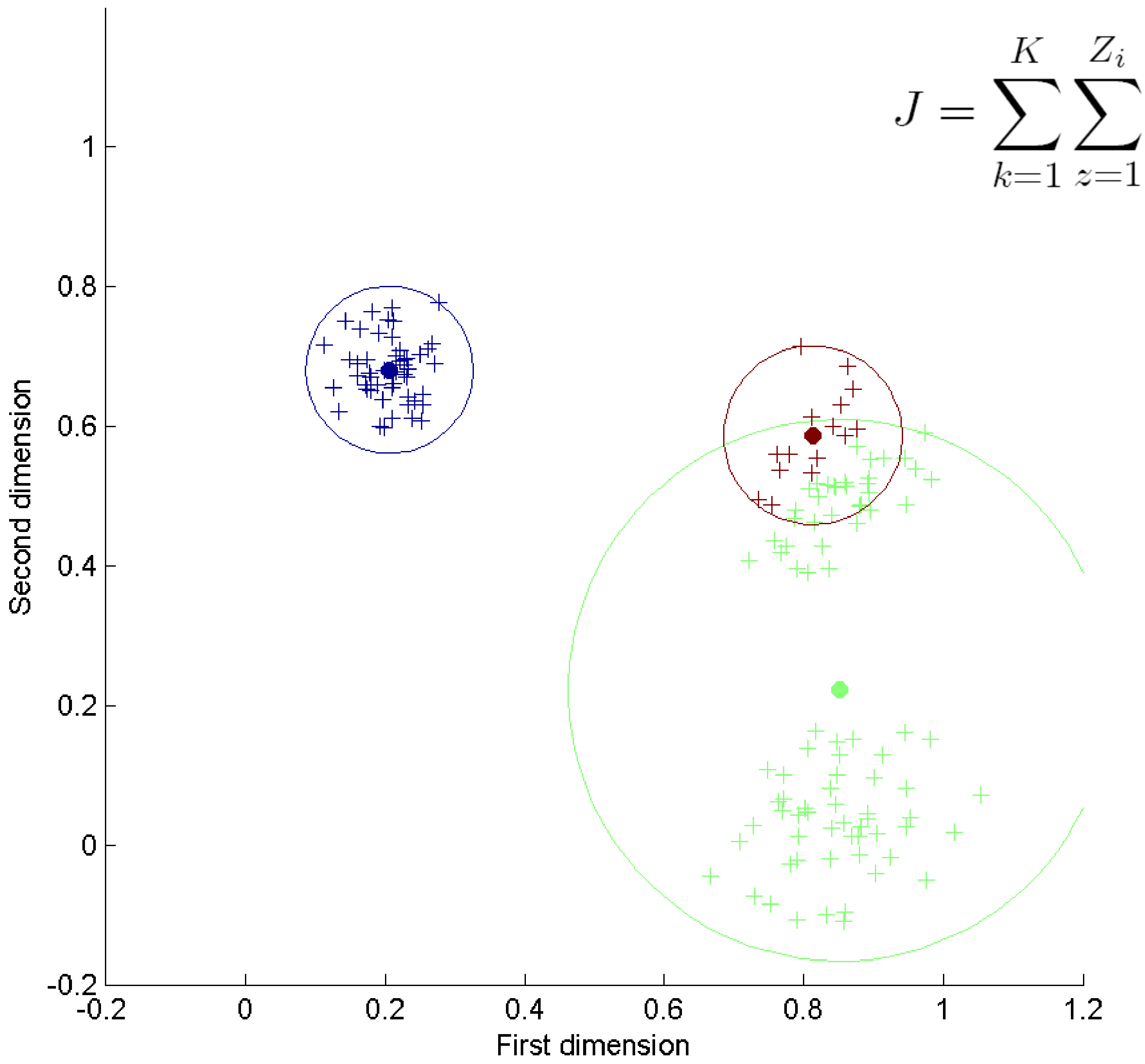


$$J = \sum_{k=1}^{K} \sum_{z=1}^{Z_i} (D_{kz} U_{kz})^2$$



k-means cluster demo. K = 3 Iteration 2. Cost J = 11.2264

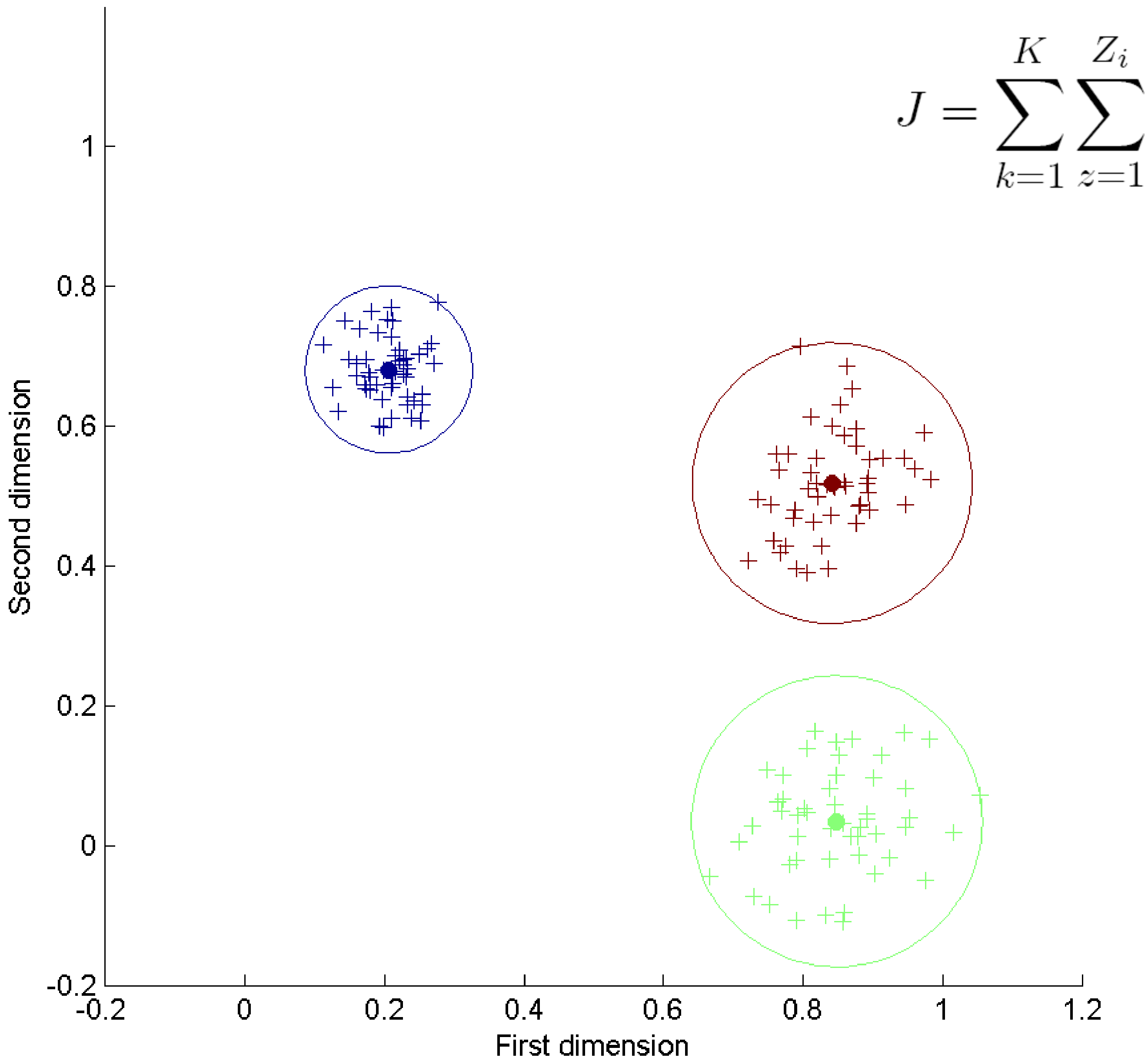
$$J = \sum_{k=1}^{K} \sum_{z=1}^{Z_i} (D_{kz} U_{kz})^2$$



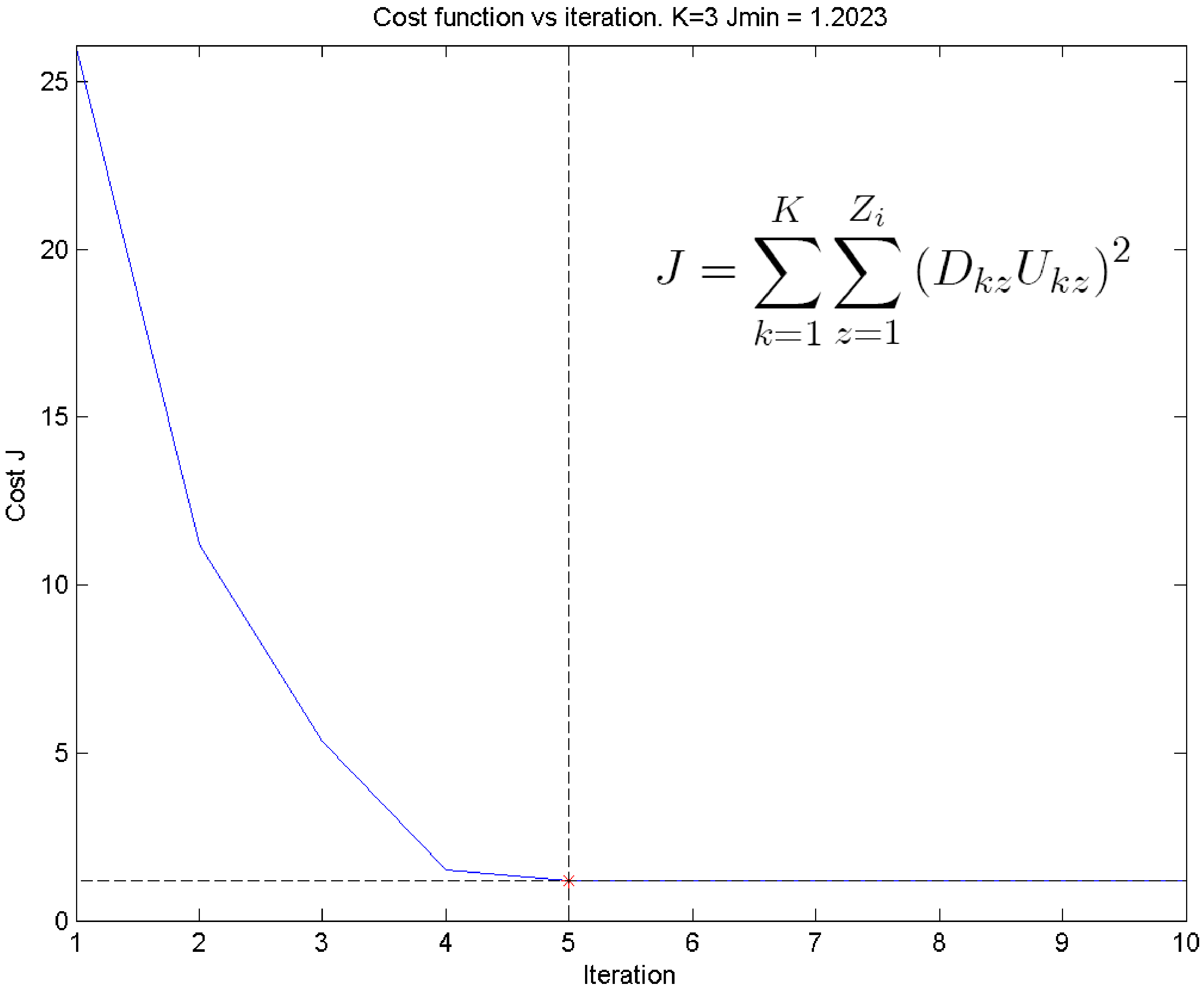
k-means cluster demo. K = 3 Iteration 3. Cost J = 5.3459

$$J = \sum_{k=1}^{K} \sum_{z=1}^{Z_i} (D_{kz} U_{kz})^2$$

k-means cluster demo. K = 3 Iteration 5. Cost J = 1.2023



$$J = \sum_{k=1}^{K} \sum_{z=1}^{Z_i} (D_{kz} U_{kz})^2$$


Cost function vs iteration. K=3 Jmin = 1.2023
$J = \sum_{k=1}^{K} \sum_{z=1}^{Z_i} (D_{kz} U_{kz})^2$
Cost J
0
5
10
15
20
25
Iteration
1
2
3
4
5
6
7
8
9
10

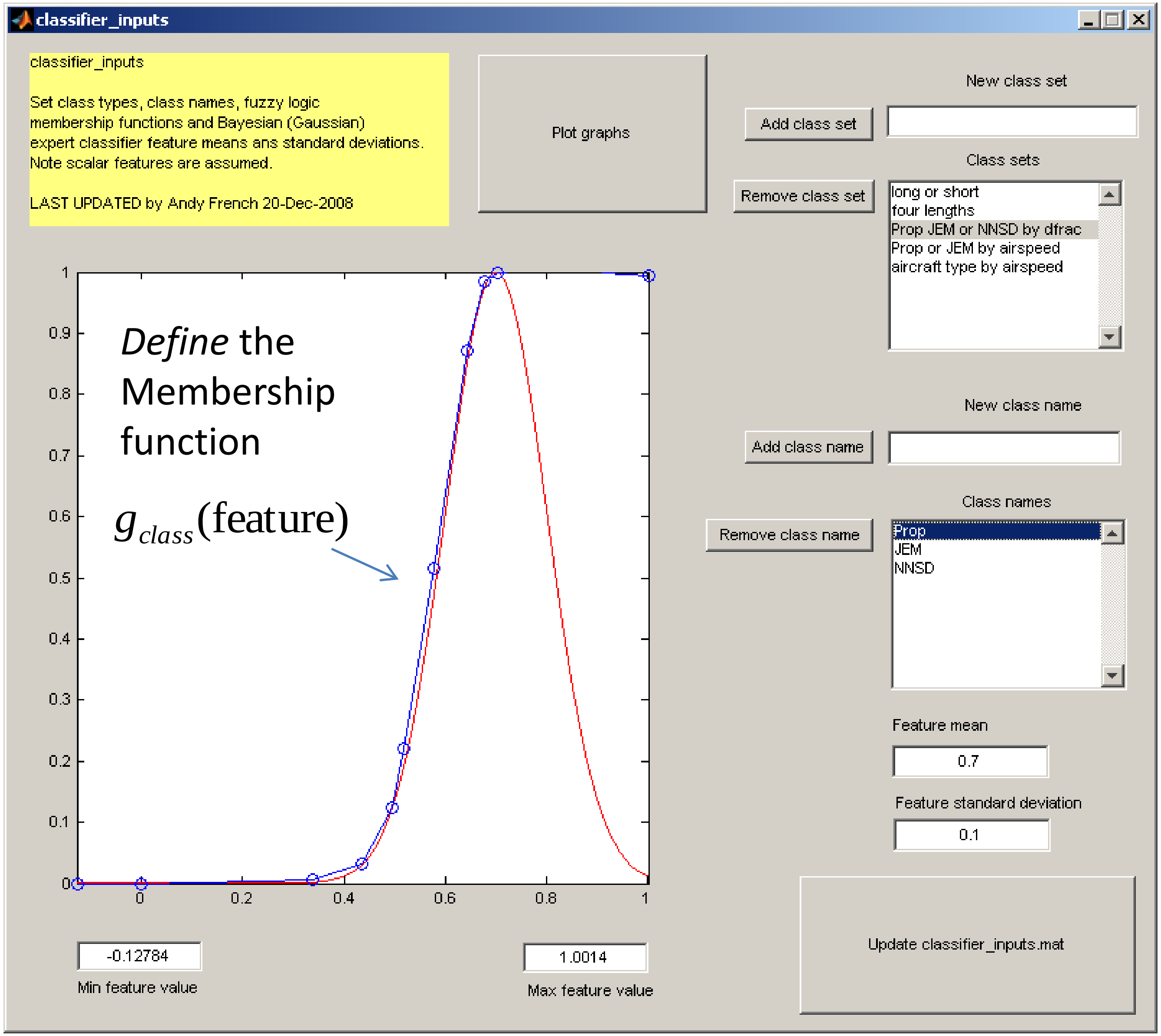
K-means classifier discriminant function

$$g_i(\mathbf{x}) = \max_k \left\{ \exp\left(-\frac{\left|\mathbf{x} - \mathbf{c}_k^{(i)}\right|^2}{2R_k^{(i)} R_k^{(i)}} \right) \right\}$$

Alternative “Fuzzy” membership matrix

$$[\mathbf{U}]_{kz} \equiv U_{kz} \rightarrow \frac{1}{\sum_{v=1}^{K} \left(\frac{D_{kz}}{D_{vz}} \right)^{\frac{2}{F-1}}}$$

Radar example: Gaussian & Fuzzy logic classification methods employed



classifier_inputs

Set class types, class names, fuzzy logic membership functions and Bayesian (Gaussian) expert classifier feature means ans standard deviations. Note scalar features are assumed.

LAST UPDATED by Andy French 20-Dec-2008

Define the Membership function

$g_{class}(\text{feature})$

Class sets:

- long or short
- four lengths
- Prop JEM or NNSD by dfrac
- Prop or JEM by airspeed
- aircraft type by airspeed

Class names:

- Prop
- JEM
- NNSD

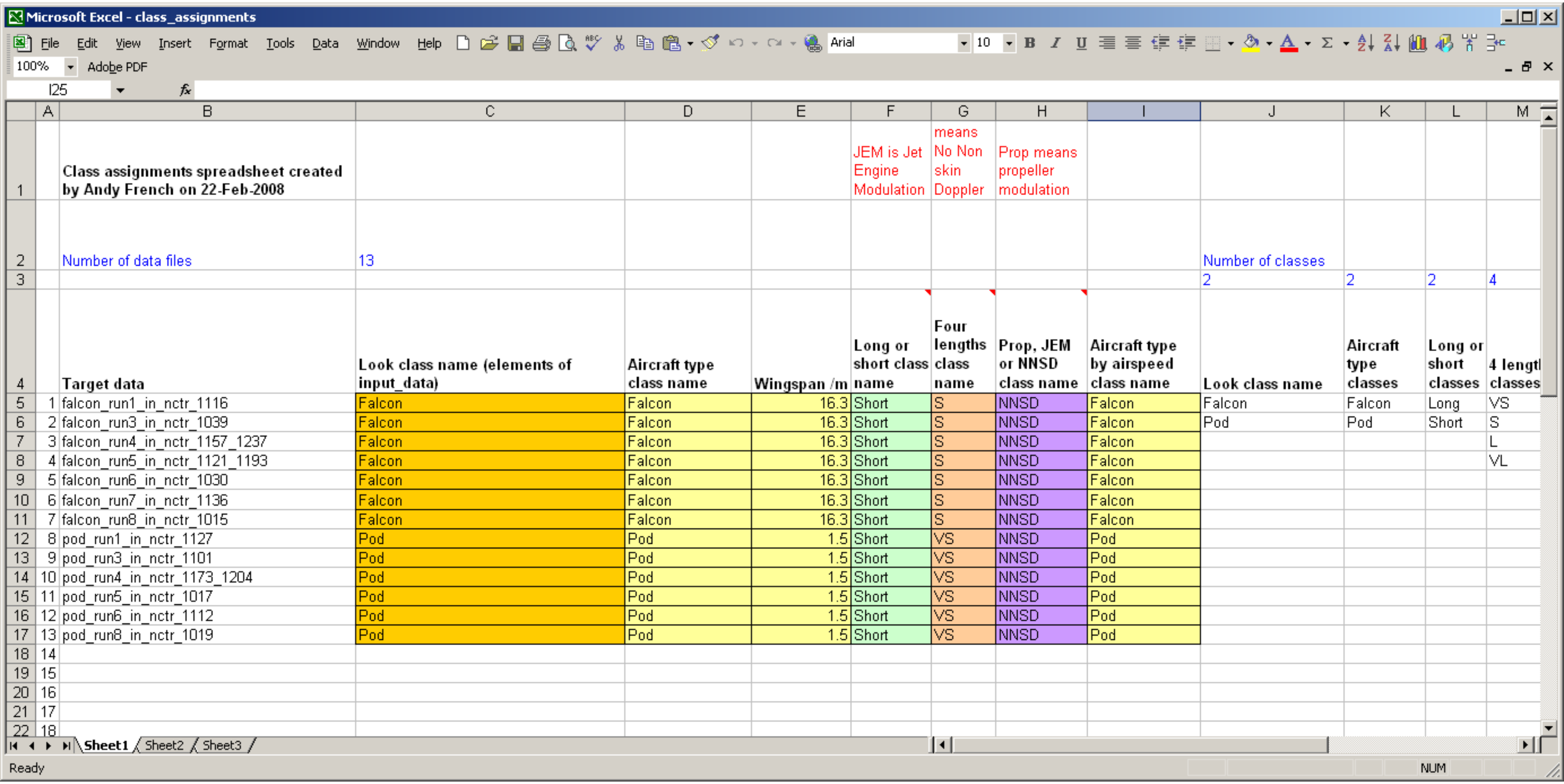
Feature mean: 0.7

Feature standard deviation: 0.1

Min feature value: -0.12784

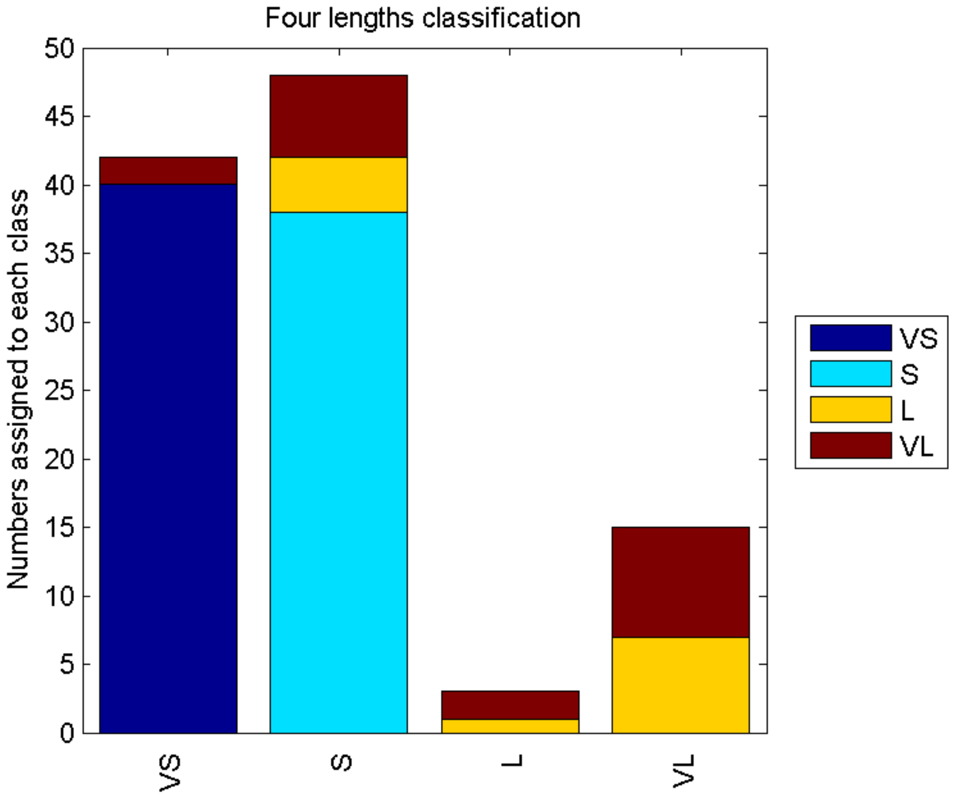
Max feature value: 1.0014

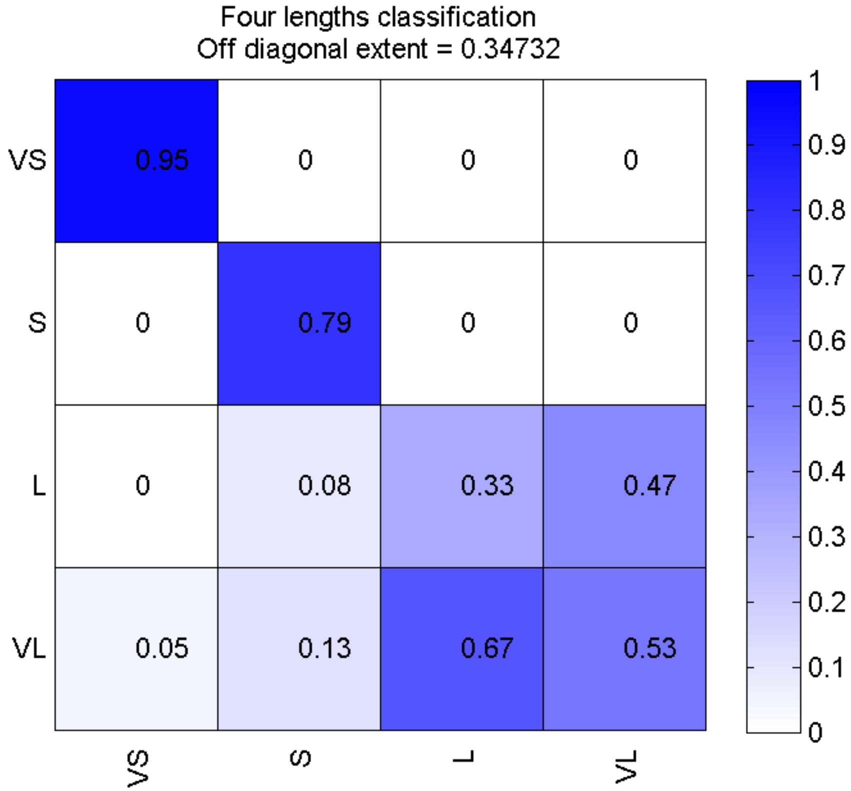
Radar target classification: truth assignments



	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Class assignments spreadsheet created by Andy French on 22-Feb-2008				JEM is Jet Engine Modulation	means No Non skin Doppler	Prop means propeller modulation					
2		Number of data files	13							Number of classes			
3										2	2	2	4
4		Target data	Look class name (elements of input_data)	Aircraft type class name	Wingspan /m	Long or short class name	Four lengths class name	Prop, JEM or NNSD class name	Aircraft type by airspeed class name	Look class name	Aircraft type classes	Long or short classes	4 lengt classes
5	1	falcon_run1_in_nctr_1116	Falcon	Falcon	16.3	Short	S	NNSD	Falcon	Falcon	Falcon	Long	VS
6	2	falcon_run3_in_nctr_1039	Falcon	Falcon	16.3	Short	S	NNSD	Falcon	Pod	Pod	Short	S
7	3	falcon_run4_in_nctr_1157_1237	Falcon	Falcon	16.3	Short	S	NNSD	Falcon				L
8	4	falcon_run5_in_nctr_1121_1193	Falcon	Falcon	16.3	Short	S	NNSD	Falcon				VL
9	5	falcon_run6_in_nctr_1030	Falcon	Falcon	16.3	Short	S	NNSD	Falcon				
10	6	falcon_run7_in_nctr_1136	Falcon	Falcon	16.3	Short	S	NNSD	Falcon				
11	7	falcon_run8_in_nctr_1015	Falcon	Falcon	16.3	Short	S	NNSD	Falcon				
12	8	pod_run1_in_nctr_1127	Pod	Pod	1.5	Short	VS	NNSD	Pod				
13	9	pod_run3_in_nctr_1101	Pod	Pod	1.5	Short	VS	NNSD	Pod				
14	10	pod_run4_in_nctr_1173_1204	Pod	Pod	1.5	Short	VS	NNSD	Pod				
15	11	pod_run5_in_nctr_1017	Pod	Pod	1.5	Short	VS	NNSD	Pod				
16	12	pod_run6_in_nctr_1112	Pod	Pod	1.5	Short	VS	NNSD	Pod				
17	13	pod_run8_in_nctr_1019	Pod	Pod	1.5	Short	VS	NNSD	Pod				
18	14												
19	15												
20	16												
21	17												
22	18												

Radar Length feature based classification


Four lengths classification


Four lengths classification
Off diagonal extent = 0.34732

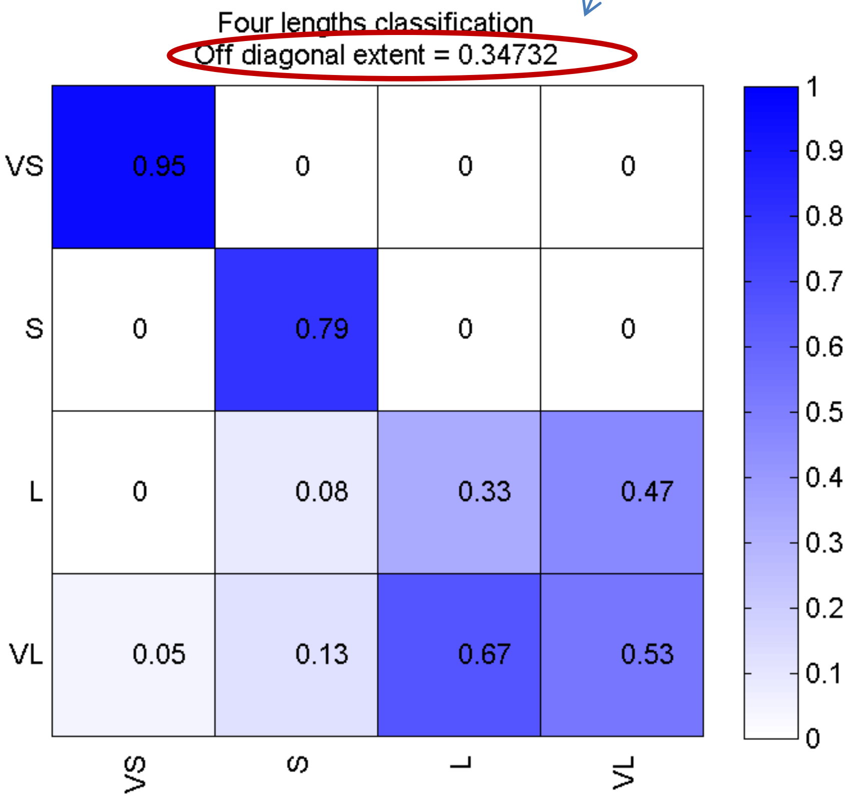
	VS	S	L	VL
VS	0.95	0	0	0
S	0	0.79	0	0
L	0	0.08	0.33	0.47
VL	0.05	0.13	0.67	0.53

Let us define a (square) *confusion matrix* $\mathbf{C}$ where element $[\mathbf{C}]_{ij} \equiv C_{ij}$ corresponds to the ratio of the number of features n_i classified as class w_i to the total number of features $m_j = \sum_i n_i$ which are actually sourced from class w_j. Perfect classification is when

$$C_{ij} = \frac{n_i}{m_j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

The *Confusion matrix* and its ‘off-diagonal-extent’

$$\xi = 1 - \frac{\mathrm{Tr}\left[\mathbf{C}\right]}{\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\mathbf{C}\right]_{ij}}$$

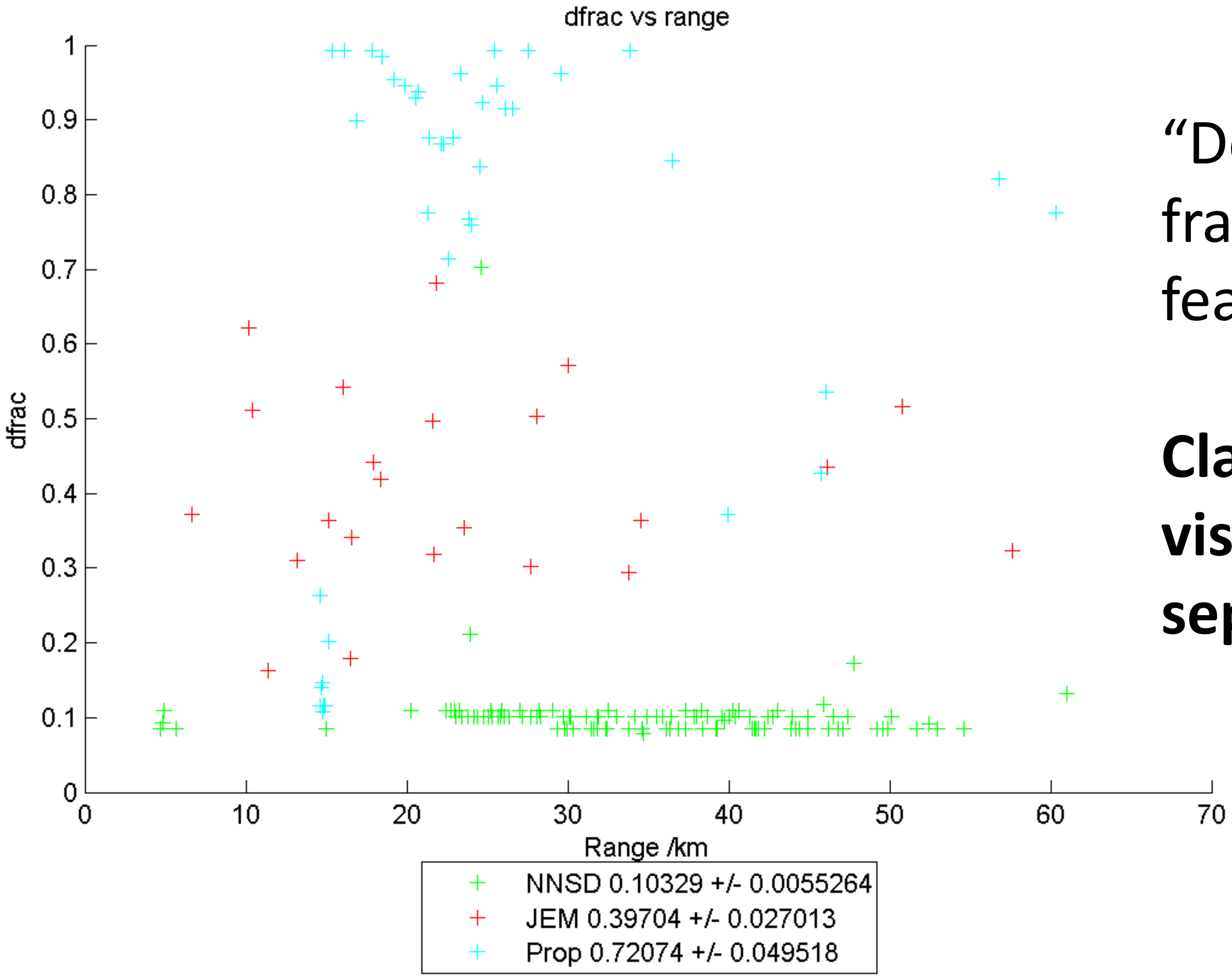


Four lengths classification
Off diagonal extent = 0.34732

	VS	S	L	VL
VS	0.95	0	0	0
S	0	0.79	0	0
L	0	0.08	0.33	0.47
VL	0.05	0.13	0.67	0.53

?

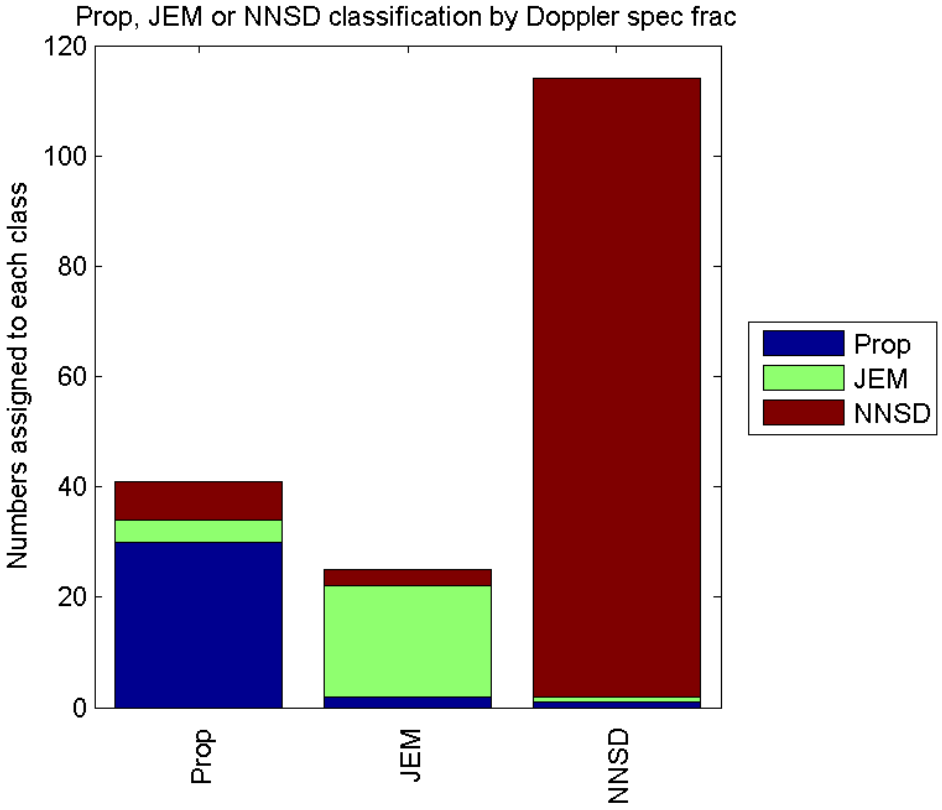
Prop, JEM or No-Non-Skin-Doppler (NNSD) classification

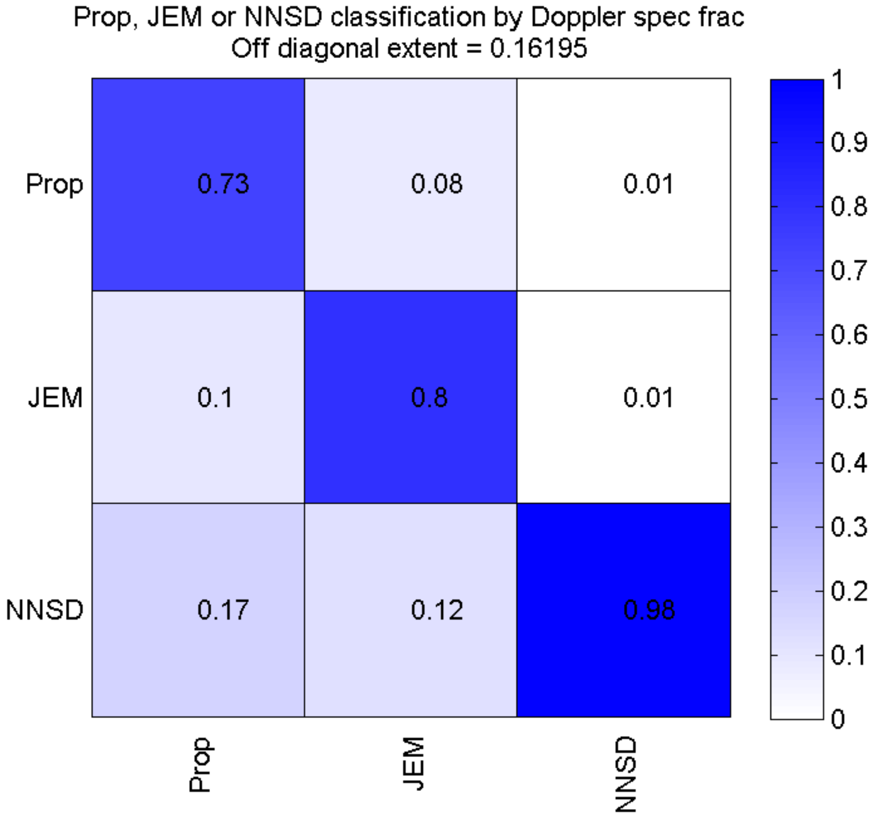


"Doppler fraction" feature

Classes are visually separable

Confusion matrix for Prop, JEM or No-Non-Skin-Doppler (NNSD) classification

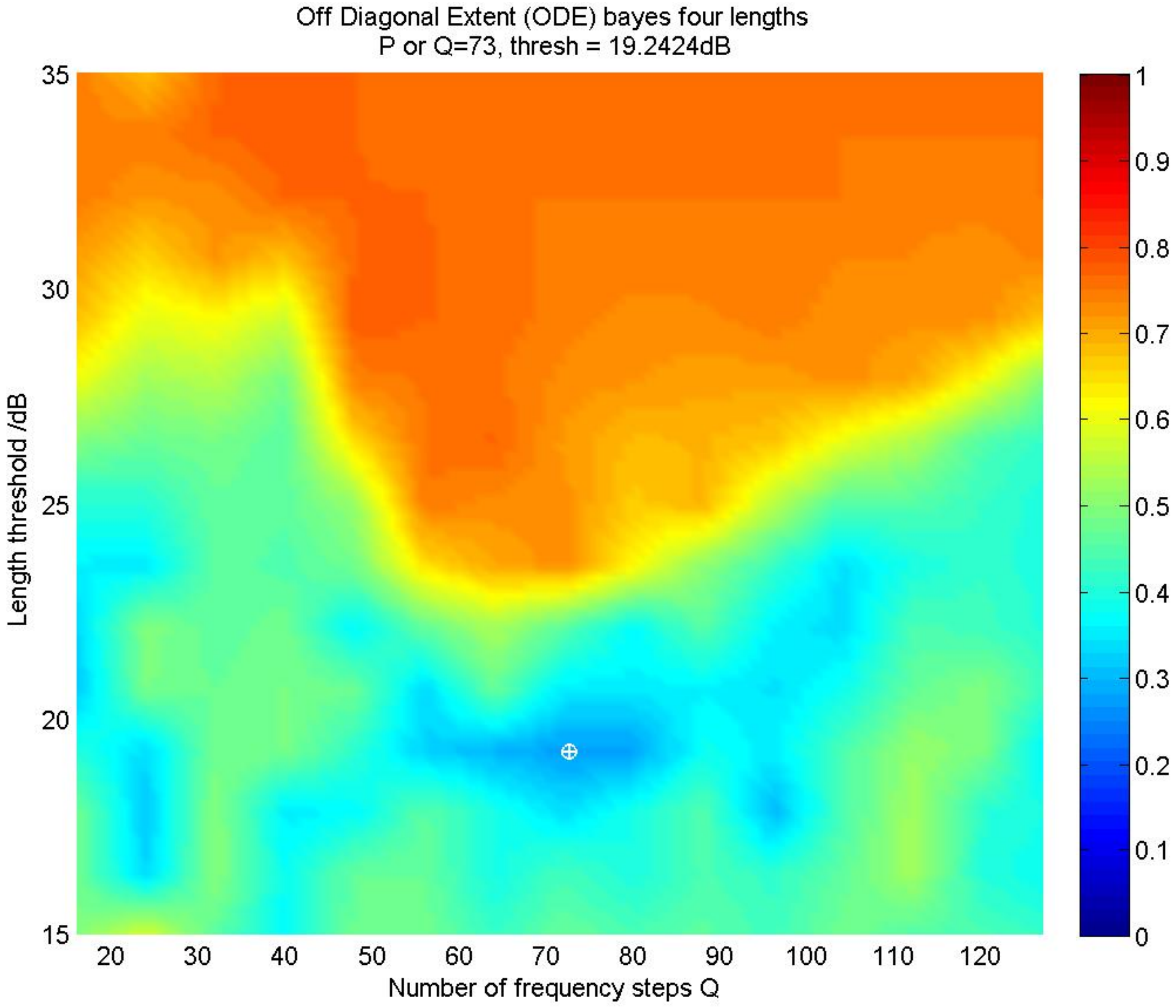



Prop, JEM or NNSD classification by Doppler spec frac
Off diagonal extent = 0.16195

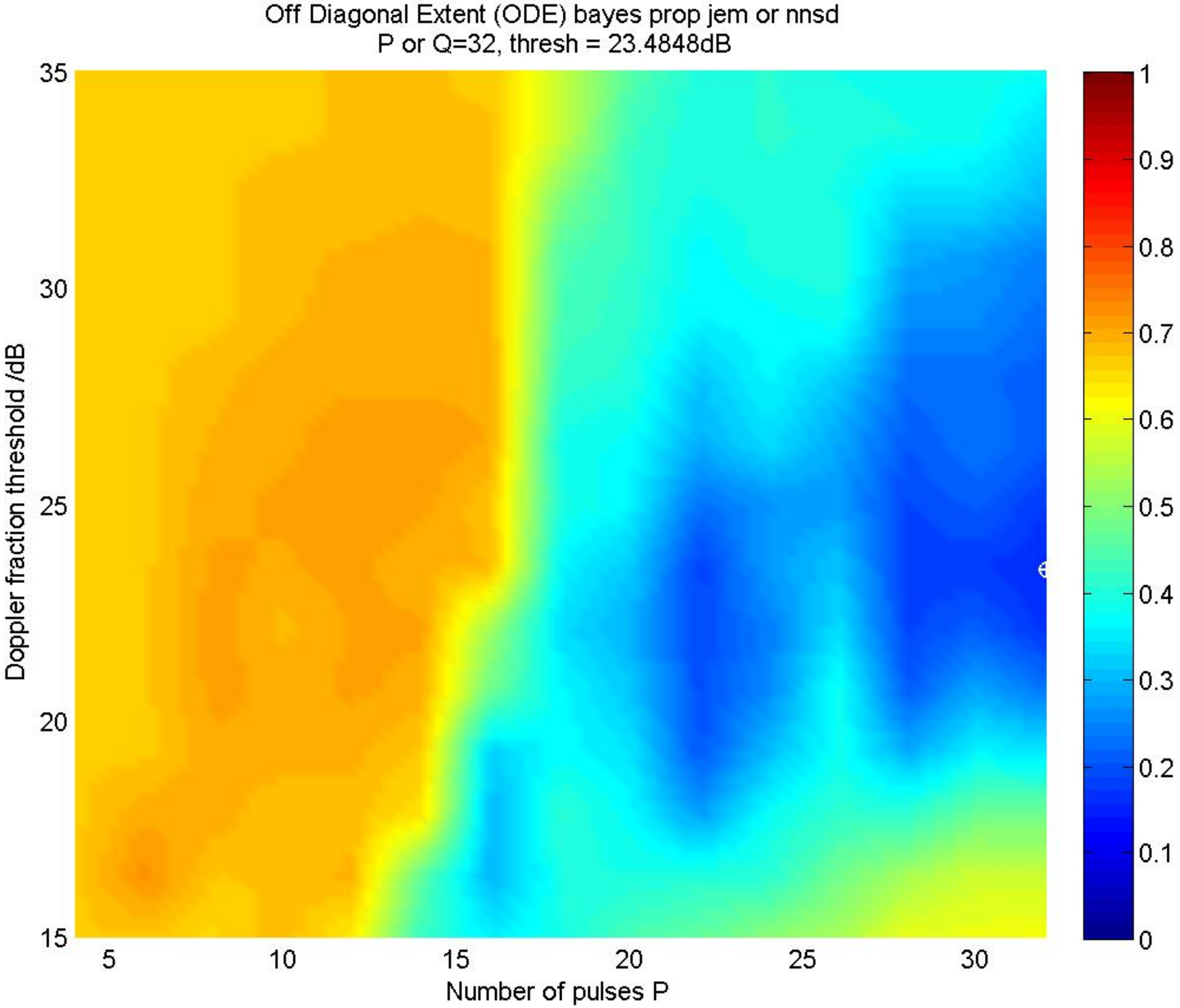
	Prop	JEM	NNSD
Prop	0.73	0.08	0.01
JEM	0.1	0.8	0.01
NNSD	0.17	0.12	0.98

Four lengths classes: VS, S, L, VL

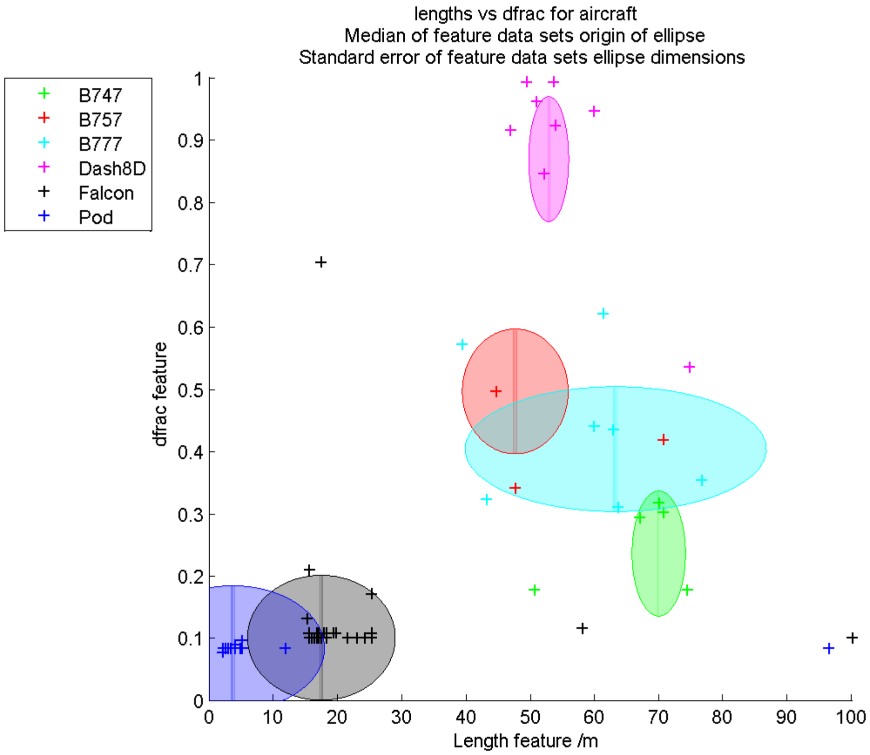
Classification performance vs length thresh & Q

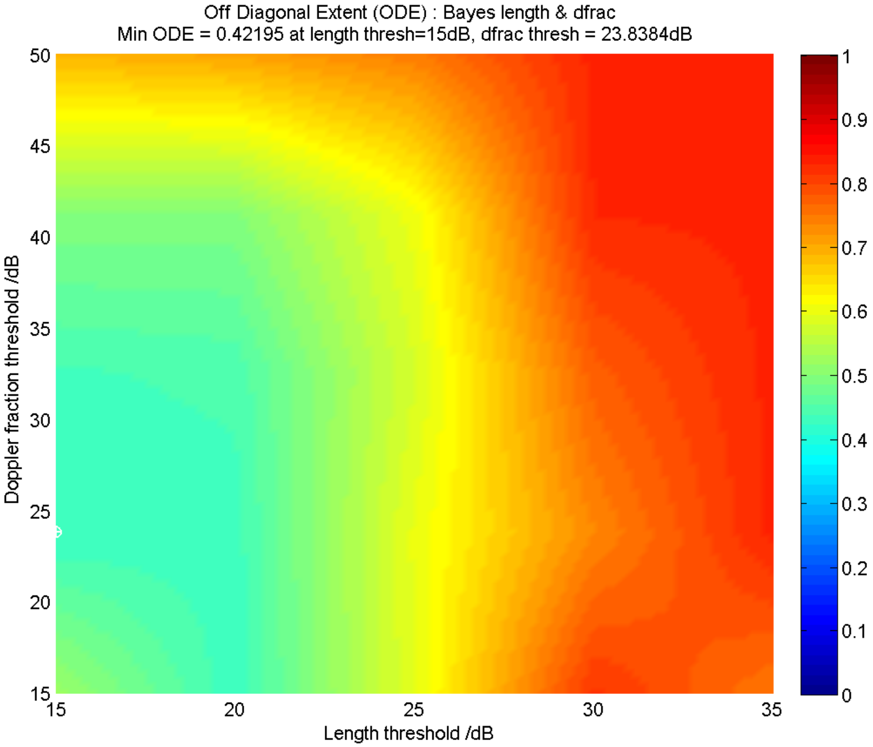


Classification performance vs dfrac thresh & P

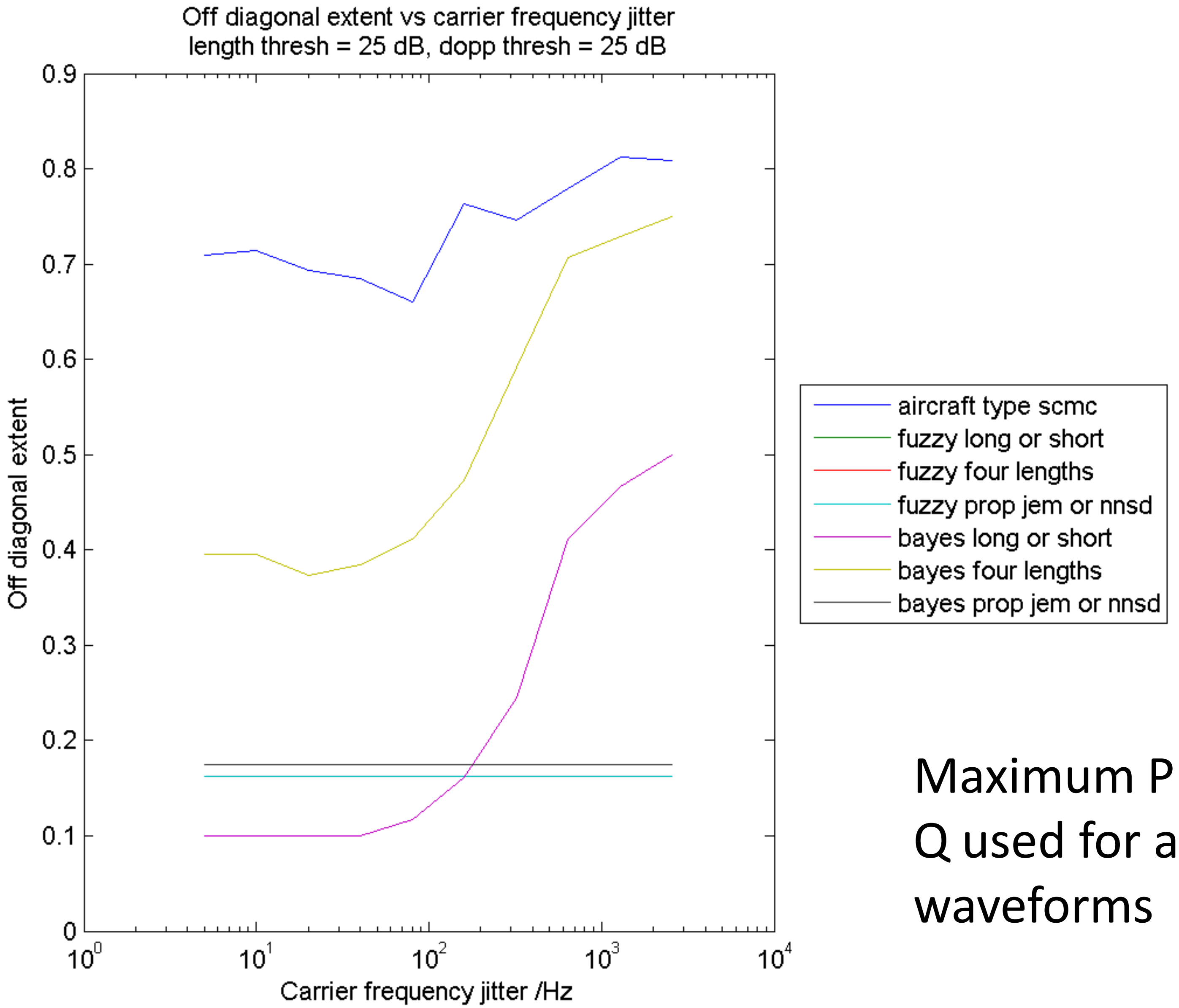


Classification based on combined length & dfrac features





Classification performance vs frequency jitter



Maximum P and Q used for all waveforms

Any questions?