

Lecture 6

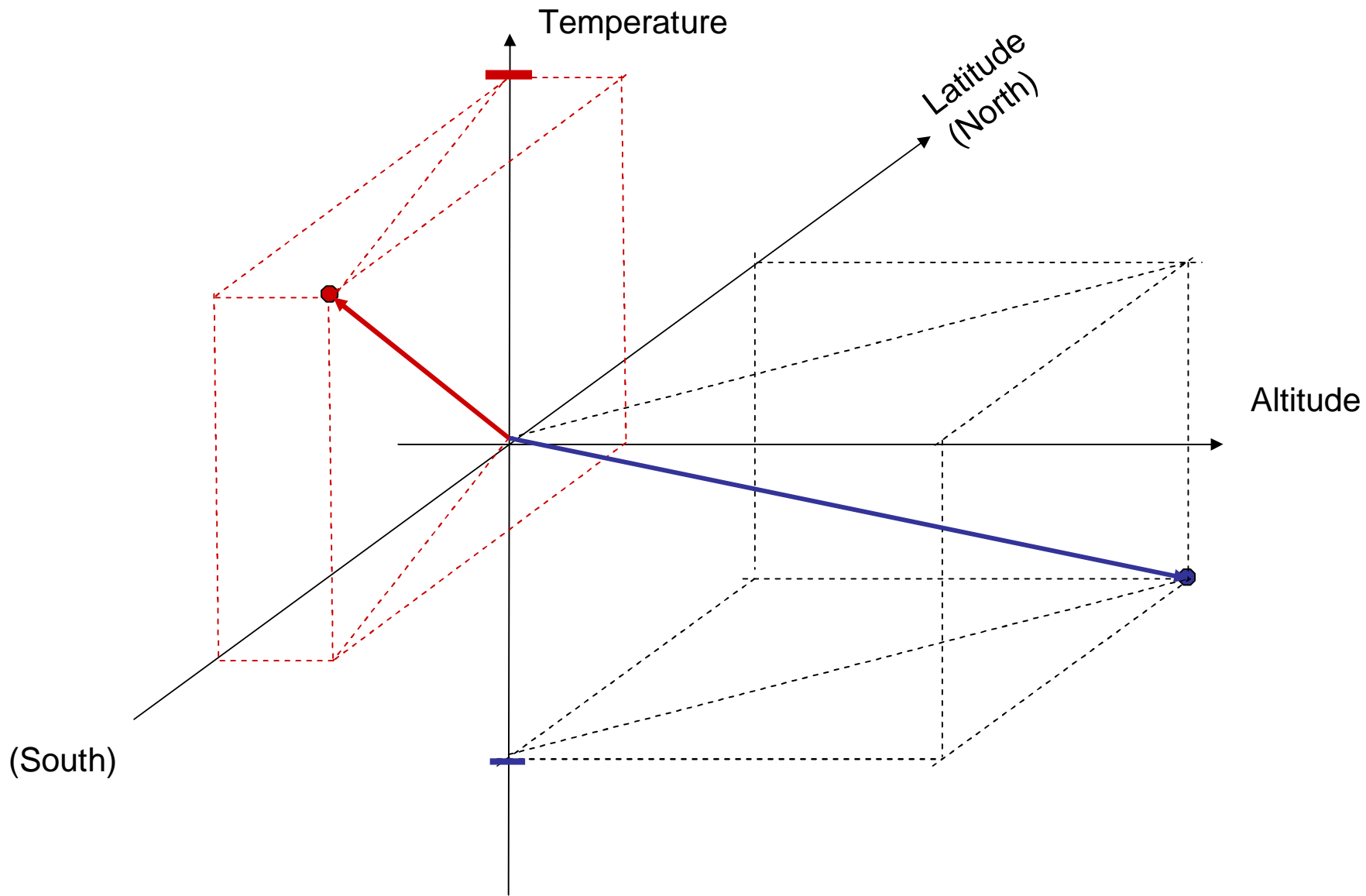
- Principal components analysis (PCA)
- A little bit about experimental design – experimental factors and their values
- Supervised analysis

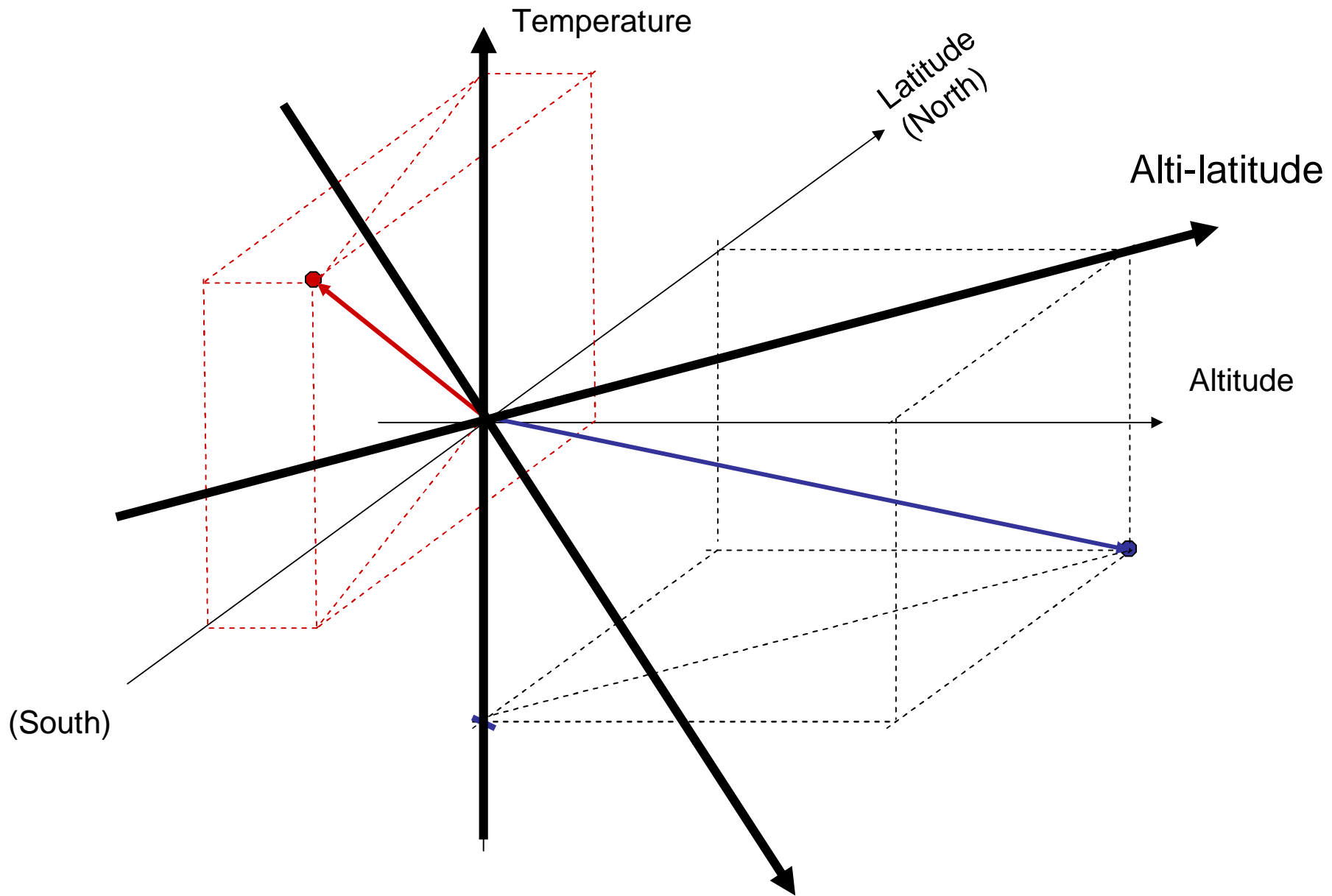
Principal Component Analysis (PCA)

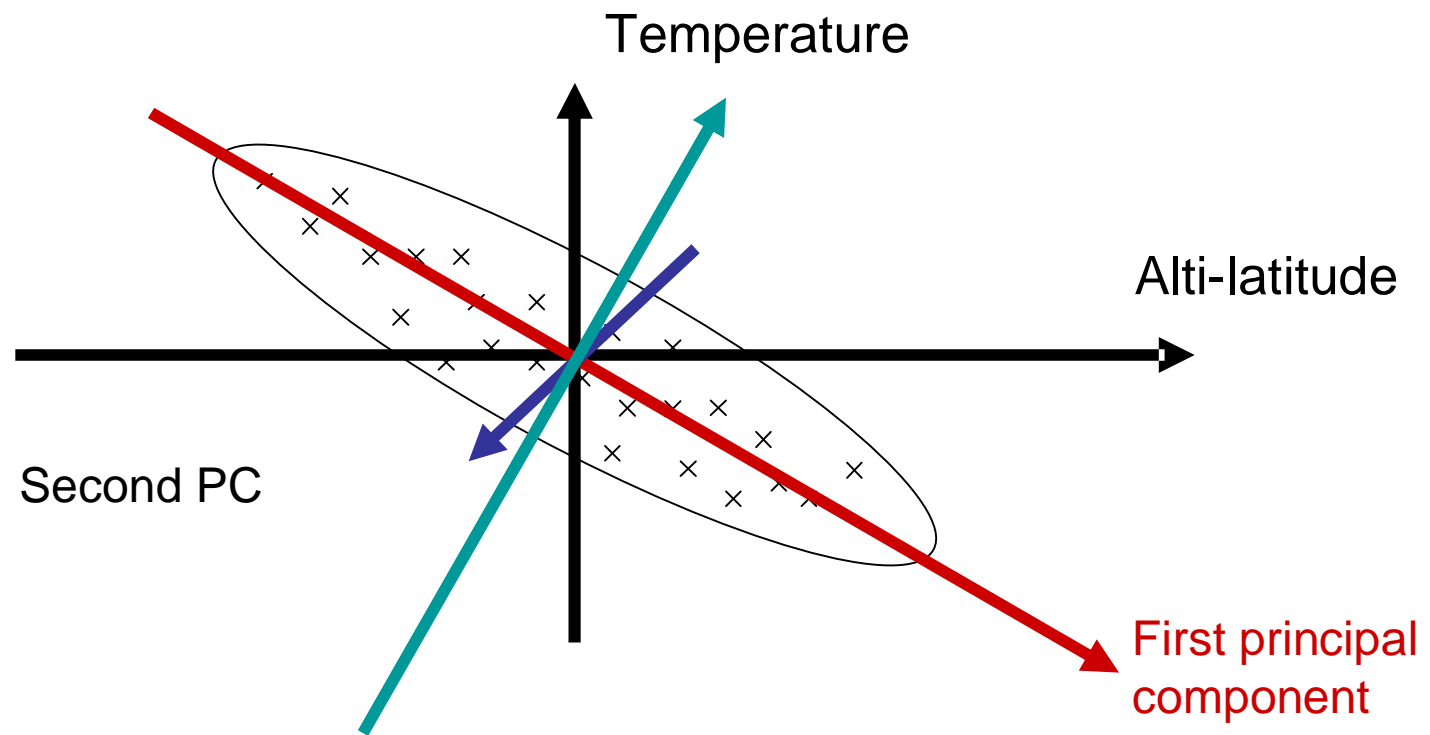
- Also known as Ordination or SVD (each version having slightly different meaning)
- Fairly nontrivial mathematical apparatus, but quite simple idea

	Condition 1	Condition 2	Condition 3
Gene 1			
Gene 2			
Gene n			

	Temperature	Altitude	Latitude
Measurement 1			
Measurement 2			
Measurement n			



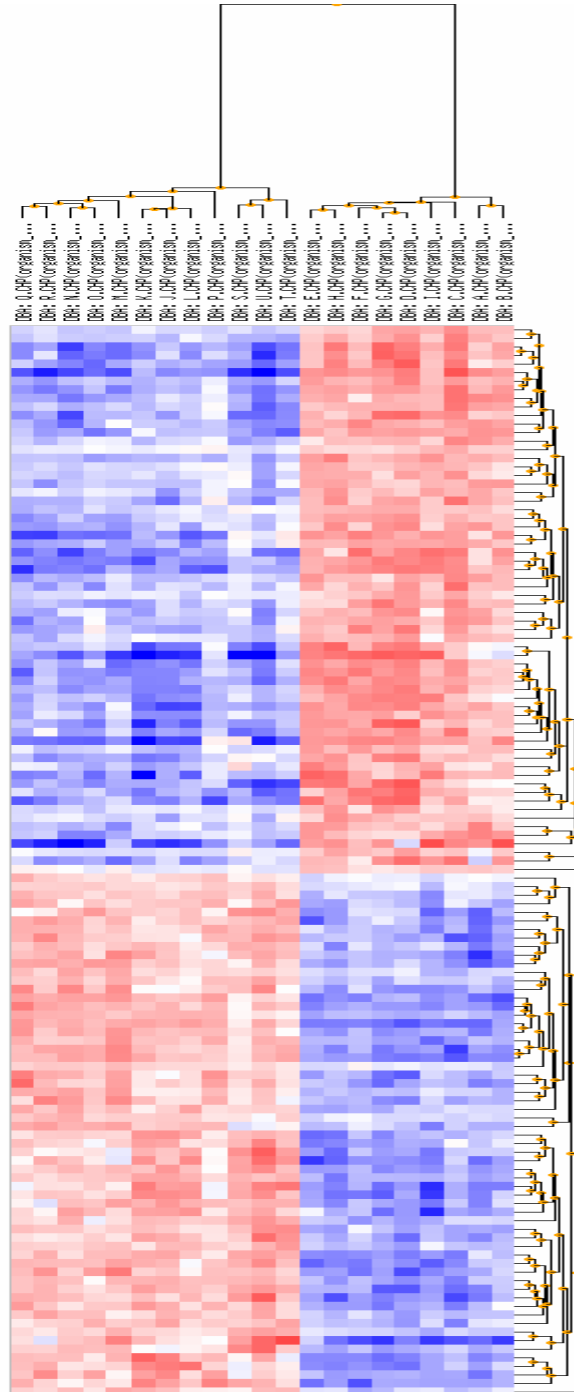


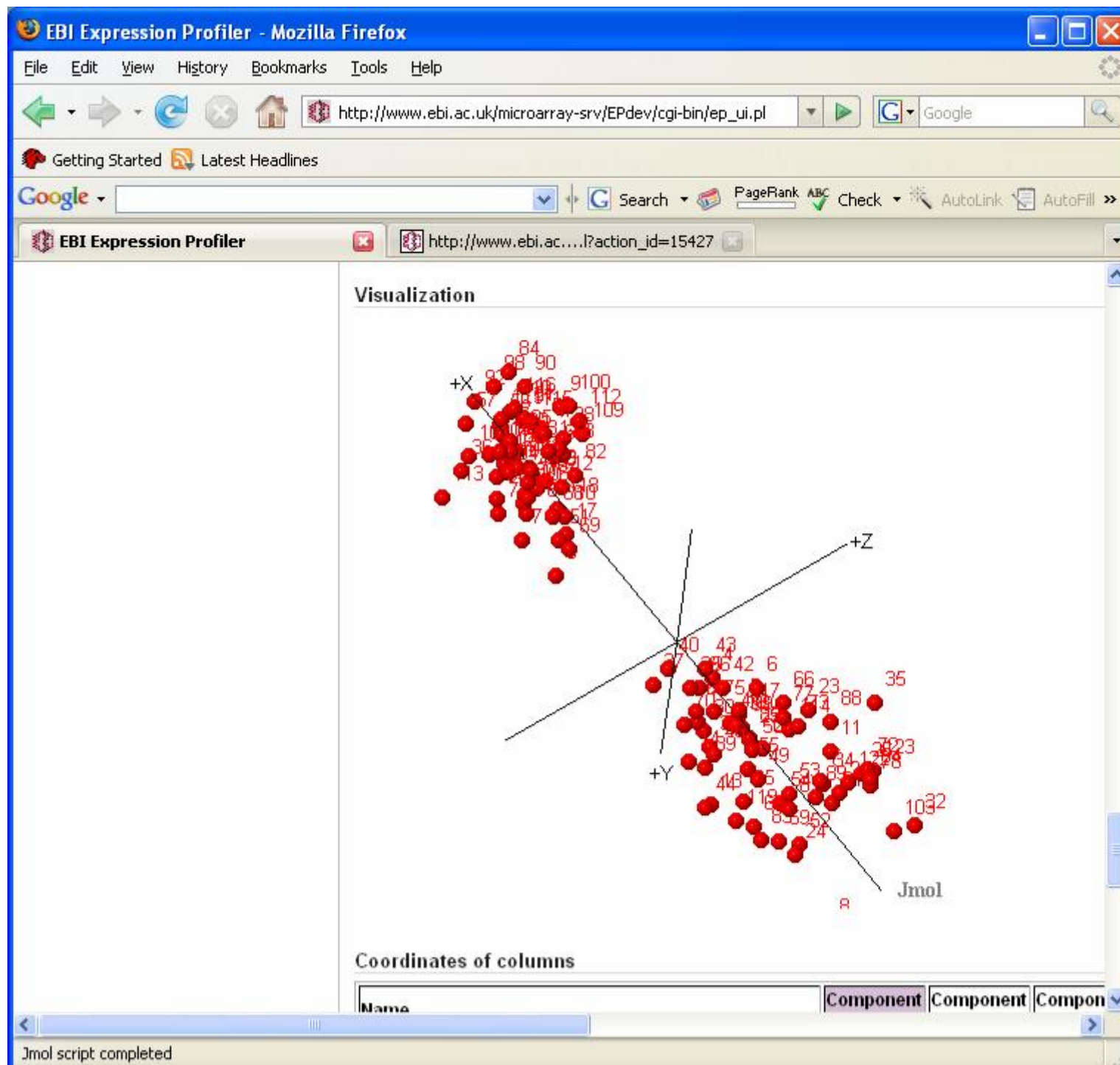


PCA in a nutshell

- The main idea – in the original n -dimensional space find the direction of most data variability (i.e., in which direction data-points are most ‘stretched’)
- Orient a new coordinate axis in this direction. This will be the first principal component, and the relative ‘stretch’ is the first eigenvalue, and the direction is the first eigenvector
- Then find the direction of the next highest variability orthogonal to the first eigenvector – this is the second component
- And so on

200069 .at; Q15020;
218139 .at; Q9H0R1;
21582 .at; Q9BV23; E...
45288 .at; Q9BV23; EN...
21679 .at; Q9BV23;
22710 .at; AAP88855; ...
21786 .at; Q724R8; E...
22754 .at; Q96Q11; E...
23584 .at; Q15006; E...
18889 .at; Q8WTT2; E...
27385 .at; Q8LY26; E...
22843 .at; Q86V18; E...
1558943 .at; AAH01610;
23893 .at; Q16394; E...
17842 .at; AAH50708; ...
29146 .at; Q15800; E...
24648 .at; Q86WP2; E...
25734 .at; AAH32540; ...
18254 .at; Q9H029; ...
19037 .at; Q9V389; E...
19913 .at; Q81XG0; ...
21298 .at; Q9H3T5; ...
14440 .at; P18440; E...
27916 .at; Q9M0T5; ...
15602 .at; AAH62363; ...
21396 .at; AAH65930; ...
21005 .at; ?
1565752 .at; AAH62363; ...
1565754 .at; AAH653; ...
235940 .at; Q8N2B1; E...
184348 .at; Q86GR2; ...
29798 .at; Q13632; E...
20992 .at; Q727T5; ...
26468 .at; Q94940; ...
2408 .at; Q94940; ...
25109 .at; Q96BK9; E...
2495 .at; CRE45819; ...
01111 .at; P55060; E...
25082 .at; CRE45951; ...
25706 .at; P11172; ...
15160 .at; P11172; ...
29983 .at; ?
23406 .at; AAH23557; ...
28869 .at; ?
22763 .at; AAH6848; ...
25881 .at; Q8N9T7; ...
26481 .at; Q8TBD9; ...
29983 .at; Q81WH4; E...
28737 .at; Q8N8N2; E...
29983 .at; P60983; ...
29983 .at; AAQ8929; ...
22666 .at; Q96919; ...
25580 .at; Q8N17; ...
29600 .at; Q15067; ...
21649 .at; Q8N0P5; ...
10145 .at; P47712; E...
28113 .at; AAH64133; ...
18722 .at; Q96EP9; ...
13186 .at; AAH56574; ...
18089 .at; Q96RV2; E...
24222 .at; P43489; ...
218799 .at; AAQ89223; ...
25510 .at; Q9HD40; ...
18239 .at; Q8N4P8; ...
20178 .at; Q9V311; E...
24093 .at; AAH0258; ...
26000 .at; ?
26000 .at; Q9NVR0; E...
23094 .at; Q15013; E...
18224 .at; Q9H9D4; ...
22623 .at; Q9UID6; ...
24777 .at; AAH64593; ...
33224 .at; ?
26484 .at; AAH14076; ...
1565689 .at; ?
23320 .at; Q96BP0; E...
23420 .at; Q9DBU2; E...
34993 .at; Q8NBW1; E...
24370 .at; Q92369; E...
19016 .at; ENSG000; ...
23657 .at; AAH62368; ...
222045 .at; BAC4523; ...
27864 .at; BAC5196; ...
209510 .at; AAH64636; ...
184036 .at; BAC23713; ...
154037 .at; BAC23713; ...
24780 .at; AAQ09533; ...
15563 .at; Q9H761; E...
24624 .at; Q9UID6; ...
24711 .at; AAH65365; ...
42343 .at; AAH63686; ...
24377 .at; Q8TBD9; ...
21235 .at; Q60466; ...
25376 .at; Q9NWL5; E...
208776 .at; Q00231; E...
1569320 .at; Q9H751; ...
35556 .at; ?
35556 .at; ?
13758 .at; AAH62437; ...
2735 .at; ?
27448 .at; Q9NWB6; E...
253015 .at; Q72530; E...
159121 .at; Q95376; ...
28480 .at; Q9POL0; E...
12241 .at; AAH62284; ...
12374 .at; Q9UK73; E...
25311 .at; AAH6489; ...
26650 .at; Q8NGM9; E...
22044 .at; BAC45238; ...
244648 .at; ?
159527 .at; ?
89948 .at; BAC45238; ...
43649 .at; Q9V311; ...
30134 .at; Q863T6; ...
29757 .at; ?
42688 .at; ?
42113 .at; ?
27373 .at; ?
28181 .at; Q96KD7; E...
42920 .at; ?
16094 .at; ?
24778 .at; ?
33360 .at; AAP35578; ...
36839 .at; ?
1555436 .at; AAH630; ...





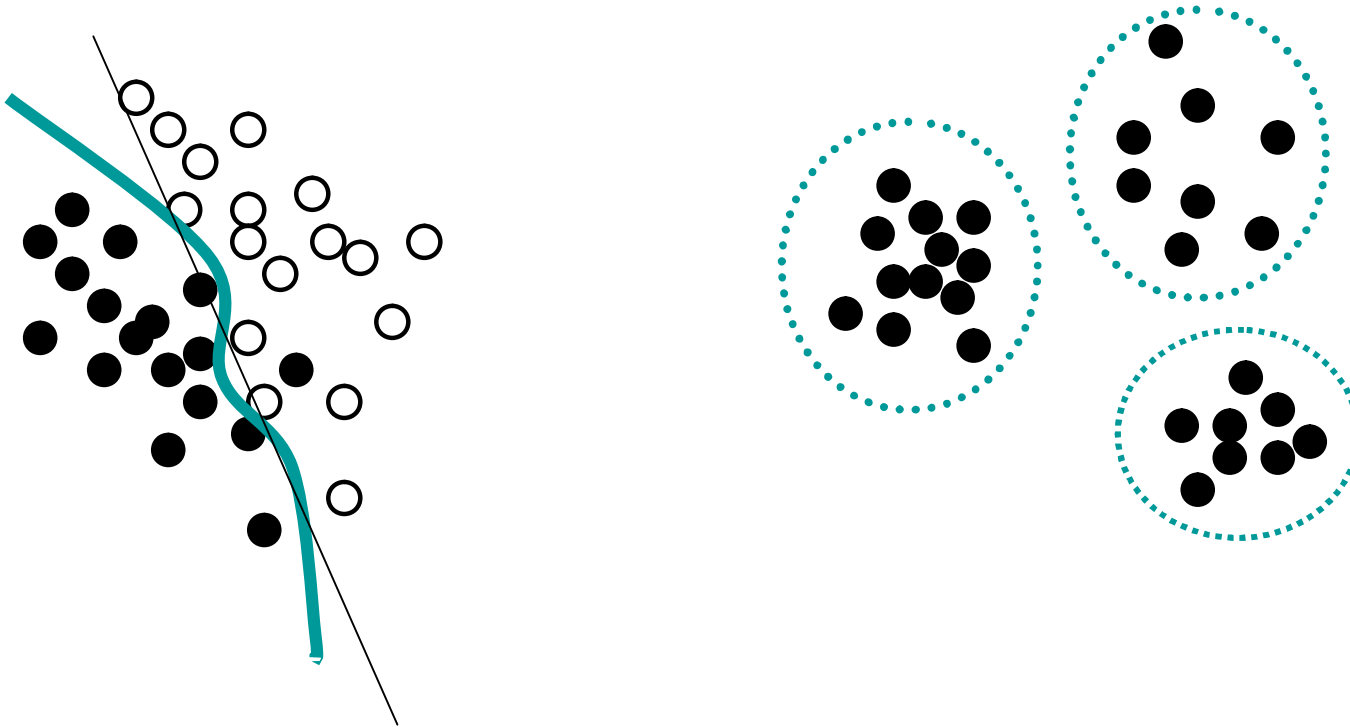
First 5
eigenvalues

1 - 17.3 (X)
2 - 0.74 (Y)
3 - 0.55 (Z)
4 - 0.44
5 - 0.37

Lecture 6

- Principal components analysis (PCA)
- **Supervised analysis**
- A little bit about experimental design
- Gene Ontology

Supervised vs unsupervised analysis



E-MEXP-758 -- BioAssayData - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Reload Home Search Favorites

Address <http://www.ebi.ac.uk/aerep/datasetselection?expid=1110518162> Go Links >>

45 Experimental conditions

select all inv	EXPERIMENTAL CONDITIONS	EXPERIMENTAL FACTORS			
		age	compound	dose	organism part
<input checked="" type="checkbox"/>	H heart 73hpf TCDD rep1 hyb1 »	73 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	heart
<input checked="" type="checkbox"/>	H heart 73hpf TCDD rep2 hyb2 »	73 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	heart
<input checked="" type="checkbox"/>	H heart 73hpf TCDD rep3 hyb3 »	73 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	heart
<input checked="" type="checkbox"/>	H heart 73hpf DMSO rep1 hyb1 »	73 h	dimethyl sulfoxide	0.1 %	heart
<input checked="" type="checkbox"/>	H heart 73hpf DMSO rep2 hyb2 »	73 h	dimethyl sulfoxide	0.1 %	heart
<input checked="" type="checkbox"/>	H heart 73hpf DMSO rep3 hyb3 »	73 h	dimethyl sulfoxide	0.1 %	heart
<input checked="" type="checkbox"/>	H heart 74hpf TCDD rep1 hyb1 »	74 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	heart
<input checked="" type="checkbox"/>	H heart 74hpf TCDD rep2 hyb2 »	74 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	heart
<input checked="" type="checkbox"/>	H heart 74hpf TCDD rep3 hyb3 »	74 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	heart
<input checked="" type="checkbox"/>	H heart 74hpf DMSO rep1 hyb1 »	74 h	dimethyl sulfoxide	0.1 %, 1 ng/ml	heart
<input checked="" type="checkbox"/>	H heart 74hpf DMSO rep2 hyb2 »	74 h	dimethyl sulfoxide	0.1 %	heart
<input checked="" type="checkbox"/>	H heart 74hpf DMSO rep3 hyb3 »	74 h	dimethyl sulfoxide	0.1 %	heart
<input checked="" type="checkbox"/>	H heart 76hpf TCDD rep1 hyb1 »	76 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	heart
<input checked="" type="checkbox"/>	H heart 76hpf TCDD rep2 hyb2 »	76 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	heart
<input checked="" type="checkbox"/>	H heart 76hpf TCDD rep3 hyb3 »	76 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	heart
<input checked="" type="checkbox"/>	H heart 76hpf DMSO rep1 hyb1 »	76 h	dimethyl sulfoxide	0.1 %	heart
<input checked="" type="checkbox"/>	H heart 76hpf DMSO rep2 hyb2 »	76 h	dimethyl sulfoxide	0.1 %	heart
<input checked="" type="checkbox"/>	H heart 76hpf DMSO rep3 hyb3 »	76 h	dimethyl sulfoxide	0.1 %	heart
<input checked="" type="checkbox"/>	H heart 84hpf TCDD rep1 hyb1 »	84 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	heart
<input checked="" type="checkbox"/>	H heart 84hpf DMSO rep1 hyb1 »	84 h	dimethyl sulfoxide	0.1 %	heart
<input checked="" type="checkbox"/>	H heart 84hpf DMSO rep3 hyb3 »	84 h	dimethyl sulfoxide	0.1 %	heart

Internet

E-MEXP-758 -- BioAssayData - Microsoft Internet Explorer

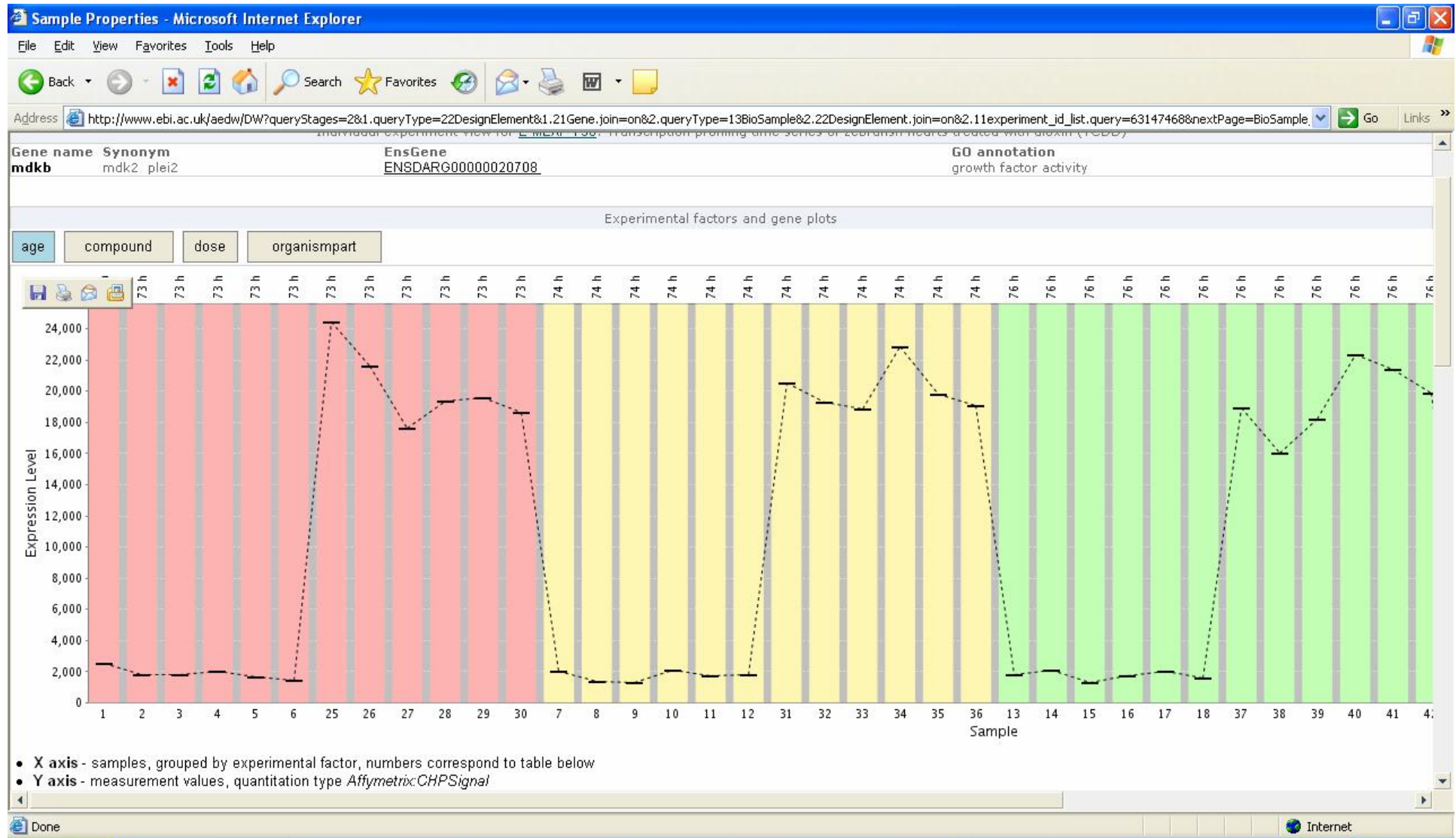
File Edit View Favorites Tools Help

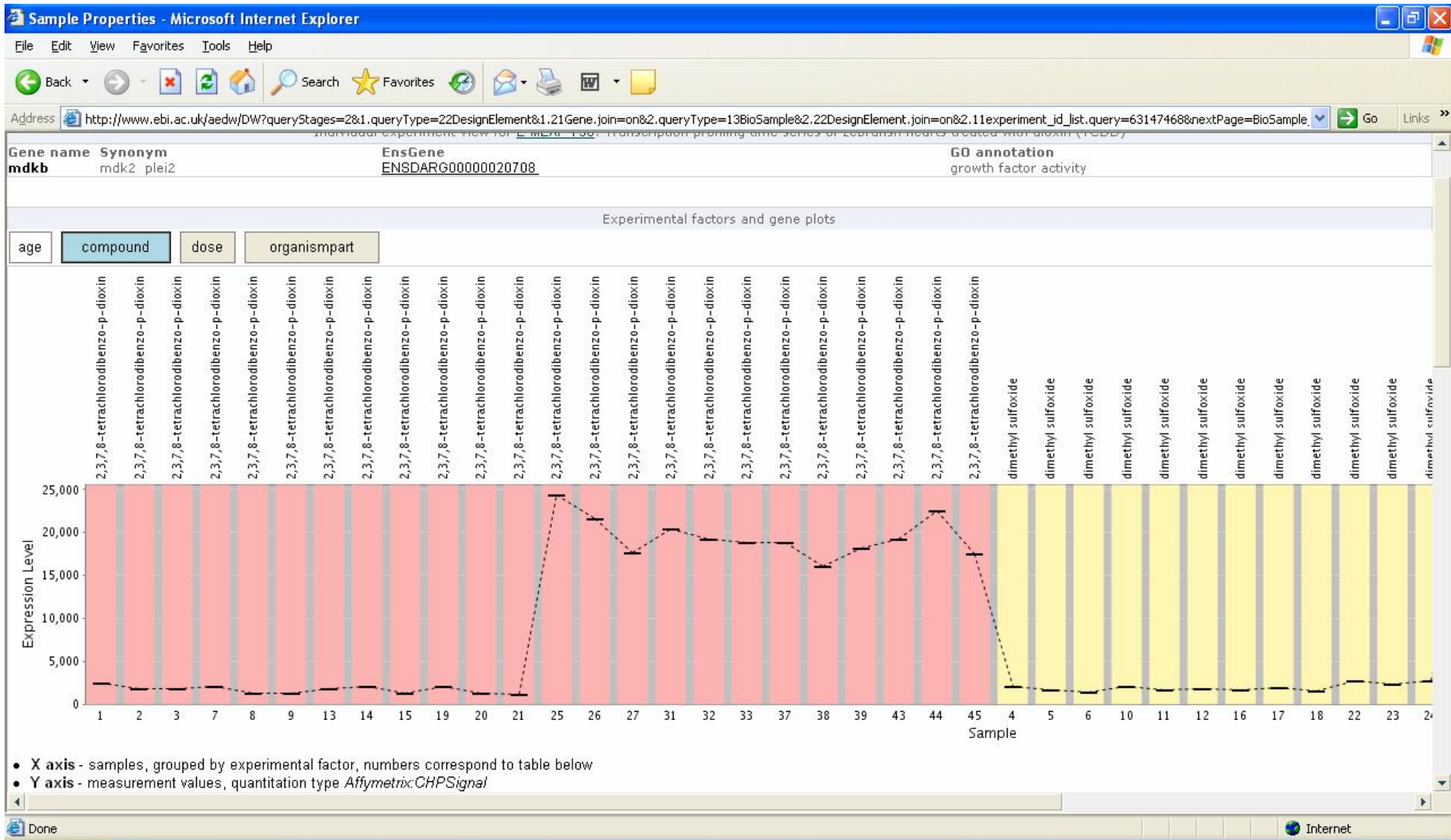
Back Forward Stop Home Search Favorites Reload Print Mail Wordpad

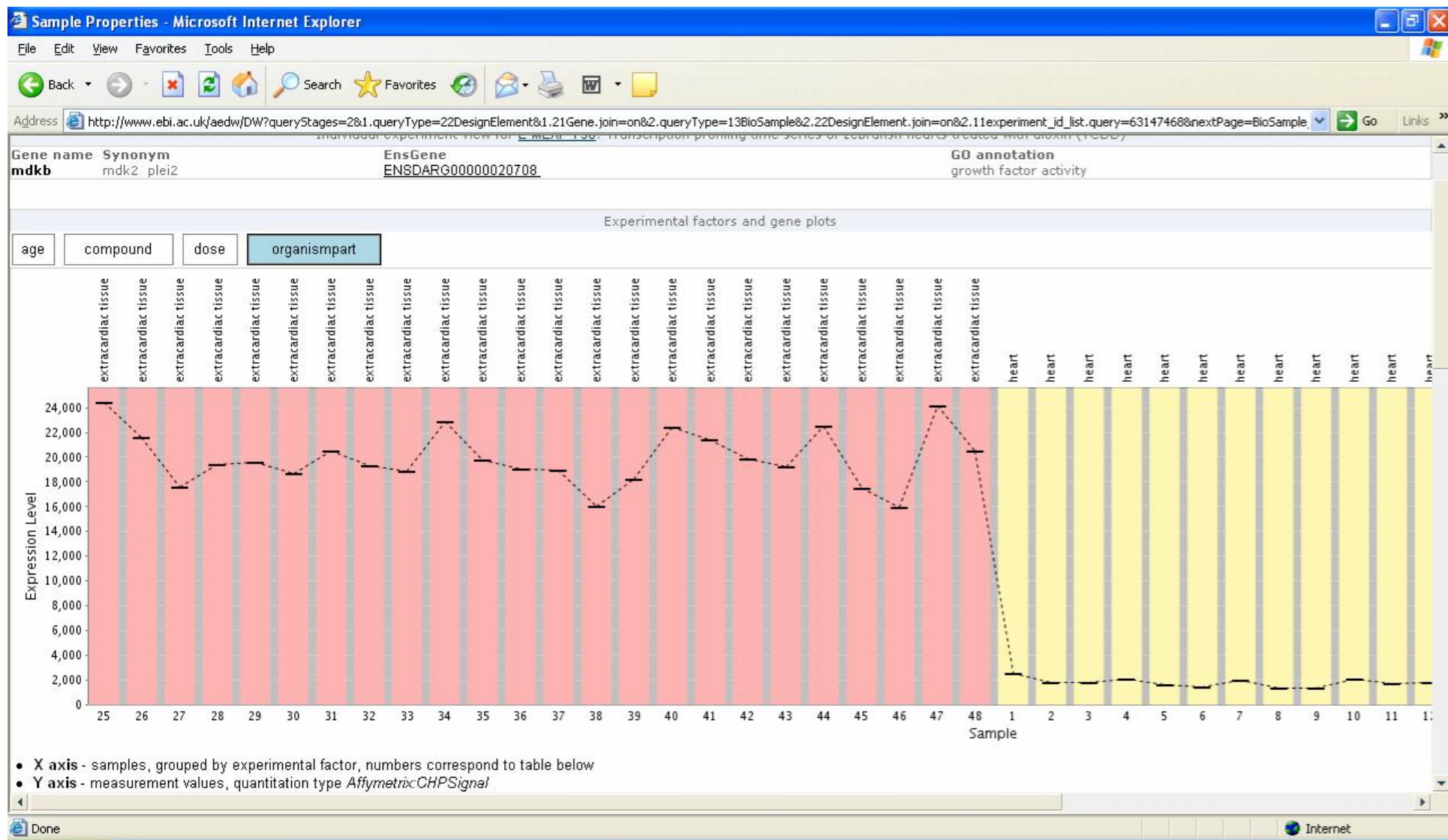
Address <http://www.ebi.ac.uk/aerep/datasetselection?expid=1110518162> Go Links

<input checked="" type="checkbox"/>	H heart 76hpf TCDD rep3 hyb3 »	76 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	heart
<input checked="" type="checkbox"/>	H heart 76hpf DMSO rep1 hyb1 »	76 h	dimethyl sulfoxide	0.1 %	heart
<input checked="" type="checkbox"/>	H heart 76hpf DMSO rep2 hyb2 »	76 h	dimethyl sulfoxide	0.1 %	heart
<input checked="" type="checkbox"/>	H heart 76hpf DMSO rep3 hyb3 »	76 h	dimethyl sulfoxide	0.1 %	heart
<input checked="" type="checkbox"/>	H heart 84hpf TCDD rep1 hyb1 »	84 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	heart
<input checked="" type="checkbox"/>	H heart 84hpf DMSO rep1 hyb1 »	84 h	dimethyl sulfoxide	0.1 %	heart
<input checked="" type="checkbox"/>	H heart 84hpf DMSO rep3 hyb3 »	84 h	dimethyl sulfoxide	0.1 %	heart
<input checked="" type="checkbox"/>	H body 73hpf TCDD rep1 hyb1 »	73 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 73hpf TCDD rep2 hyb2 »	73 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 73hpf TCDD rep3 hyb3 »	73 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 73hpf DMSO rep1 hyb1 »	73 h	dimethyl sulfoxide	0.1 %	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 73hpf DMSO rep2 hyb2 »	73 h	dimethyl sulfoxide	0.1 %	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 73hpf DMSO rep3 hyb3 »	73 h	dimethyl sulfoxide	0.1 %	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 74hpf TCDD rep1 hyb1 »	74 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 74hpf TCDD rep2 hyb2 »	74 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 74hpf TCDD rep3 hyb3 »	74 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 74hpf DMSO rep1 hyb1 »	74 h	dimethyl sulfoxide	0.1 %	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 74hpf DMSO rep2 hyb2 »	74 h	dimethyl sulfoxide	0.1 %	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 74hpf DMSO rep3 hyb3 »	74 h	dimethyl sulfoxide	0.1 %	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 76hpf TCDD rep1 hyb1 »	76 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 76hpf TCDD rep2 hyb2 »	76 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 76hpf TCDD rep3 hyb3 »	76 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 76hpf DMSO rep1 hyb1 »	76 h	dimethyl sulfoxide	0.1 %	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 76hpf DMSO rep2 hyb2 »	76 h	dimethyl sulfoxide	0.1 %	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 76hpf DMSO rep3 hyb3 »	76 h	dimethyl sulfoxide	0.1 %	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 84hpf TCDD rep1 hyb1 »	84 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 84hpf TCDD rep2 hyb2 »	84 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	extracardiac body tissue
<input checked="" type="checkbox"/>	H body 84hpf TCDD rep3 hyb3 »	84 h	2,3,7,8-tetrachlorodibenzo-p-dioxin	1 ng/ml	extracardiac body tissue

Internet



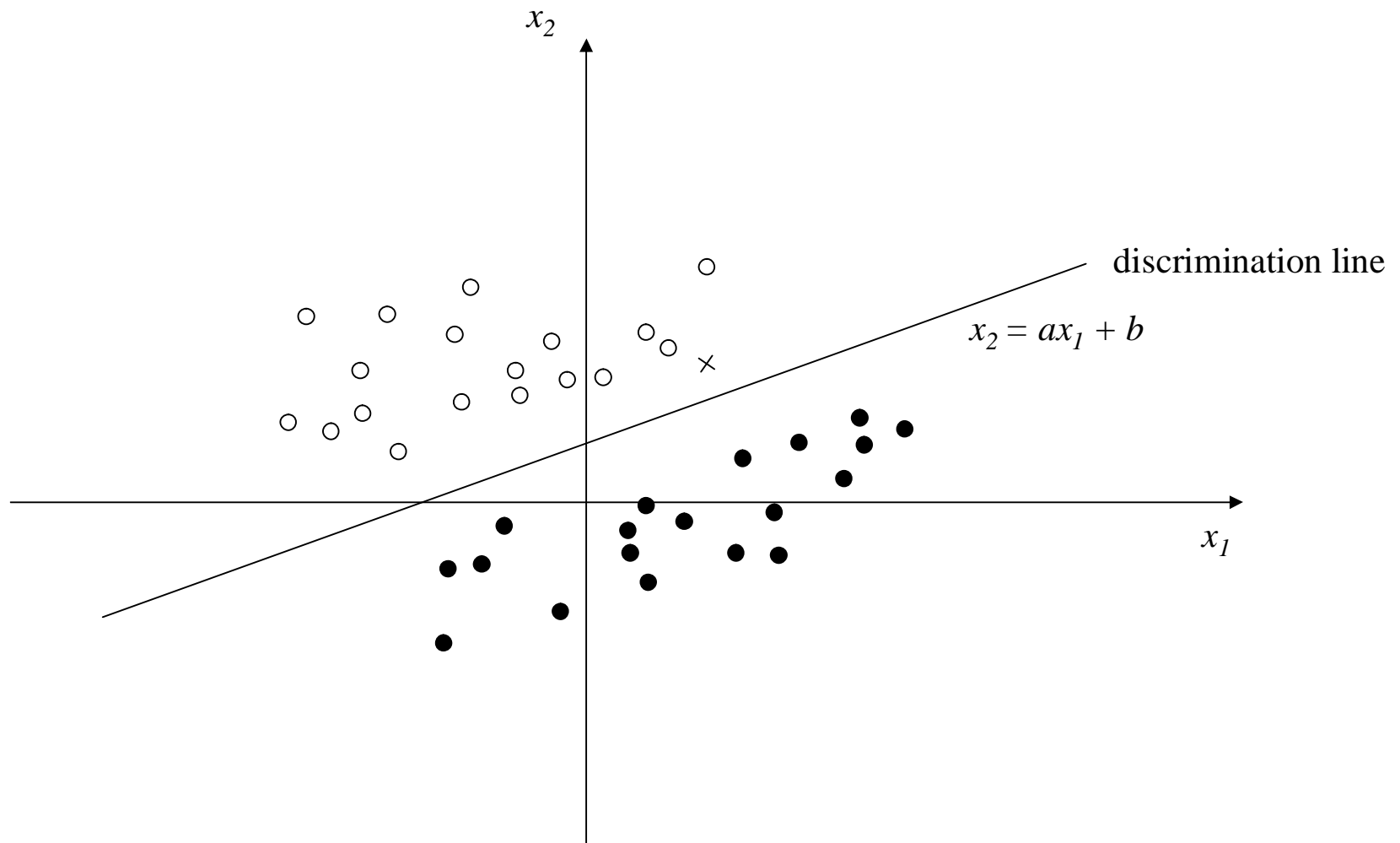


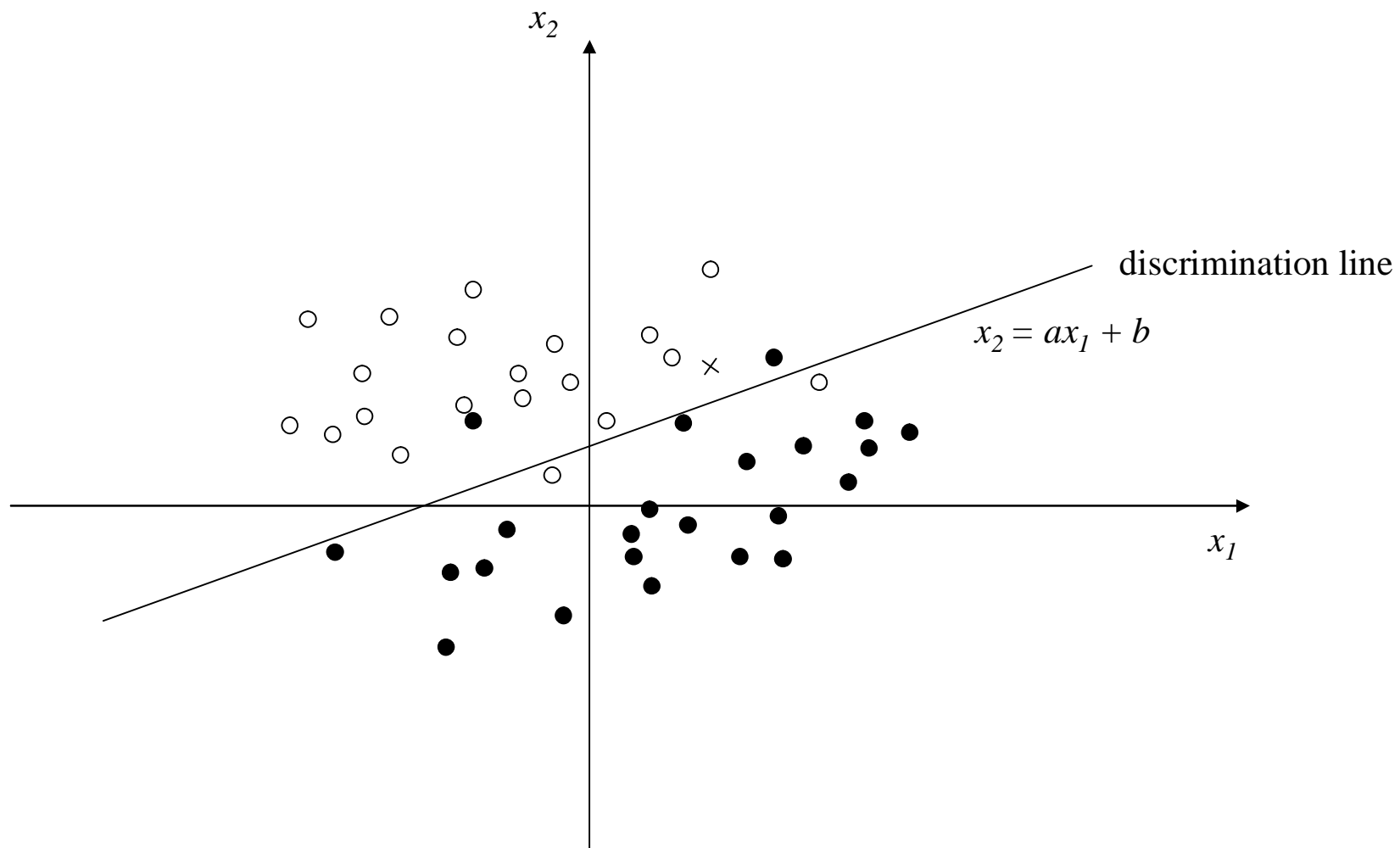


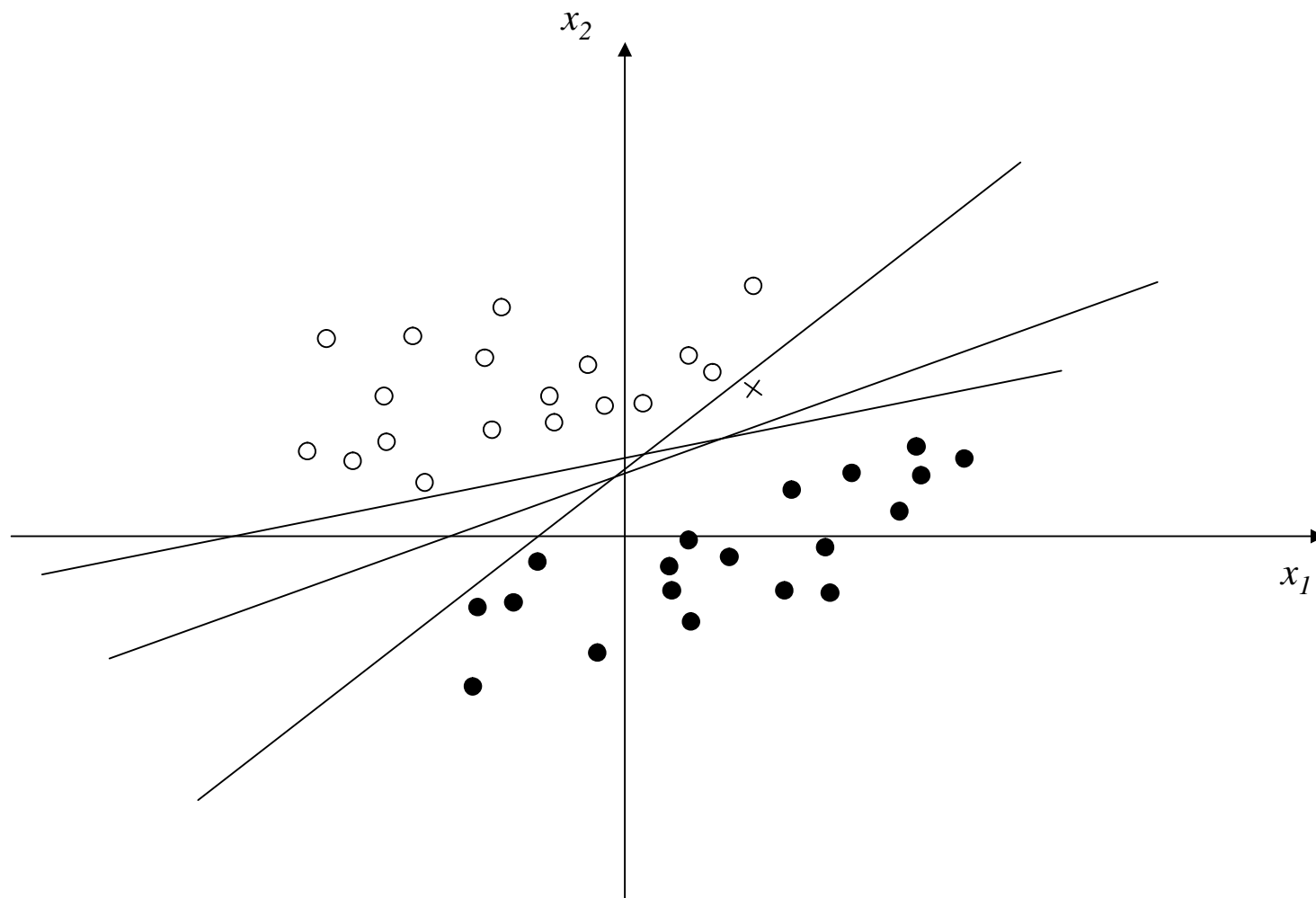
Classifiers - applications

- Training on known data – find a classifier that can separate one experimental factor value from the other based only on data
- Apply to new data – this will tell us where the new sample belongs (e.g., diseased or normal – diagnostics)

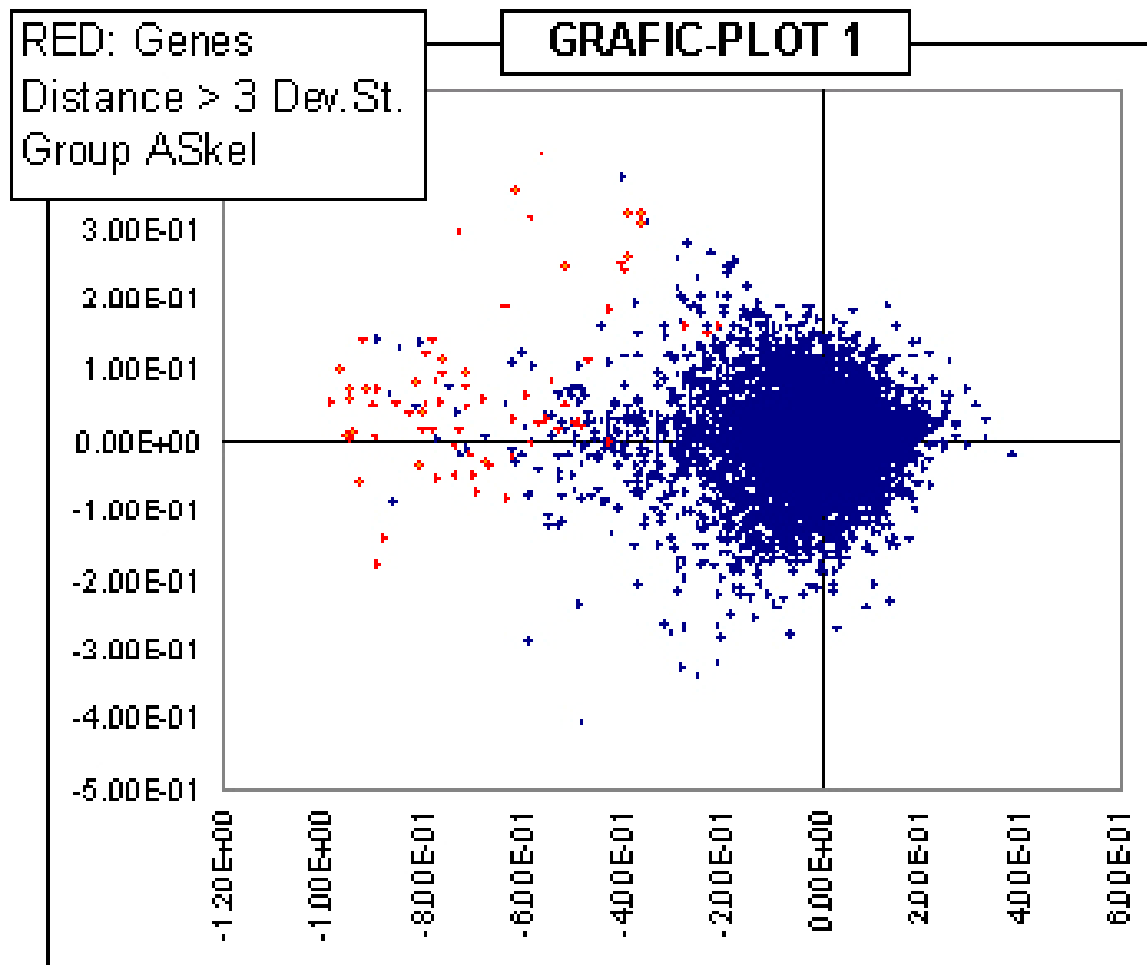
Linear discriminants



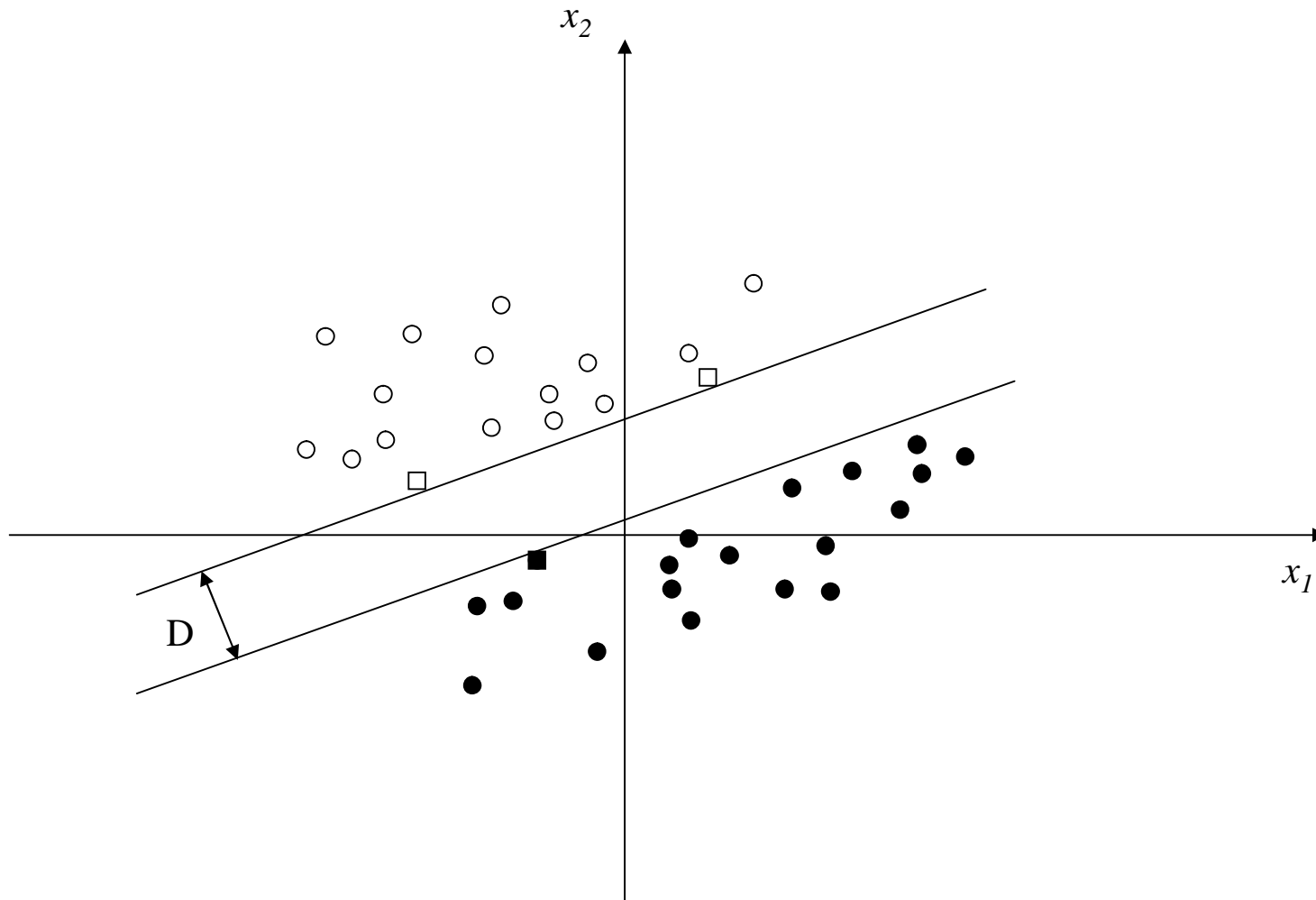


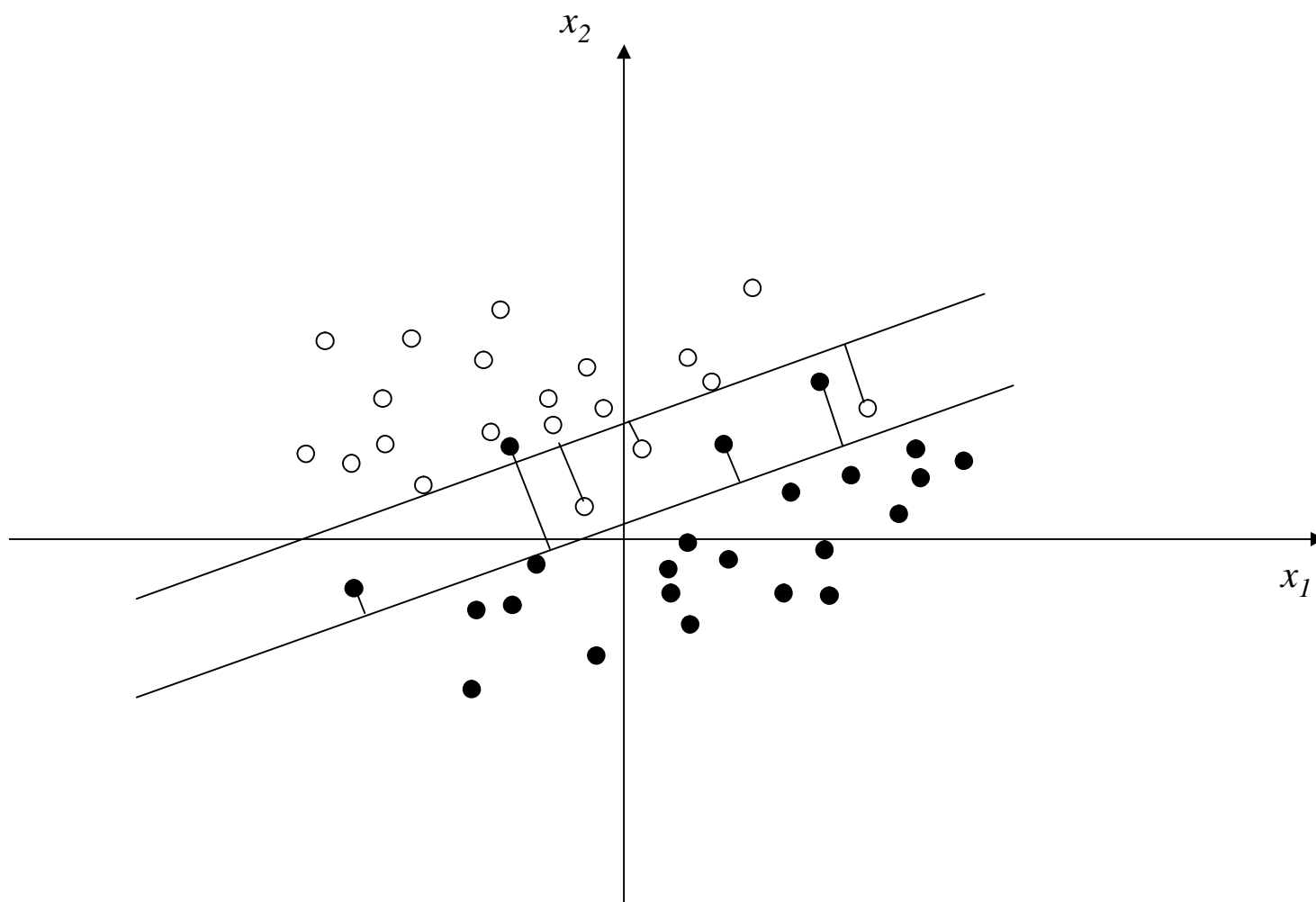


In between group analysis

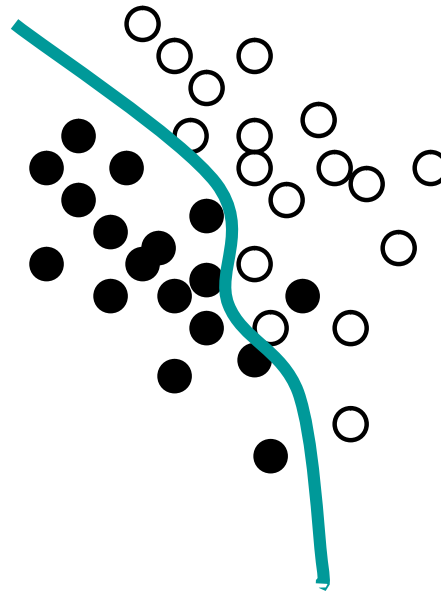


Support vector machines





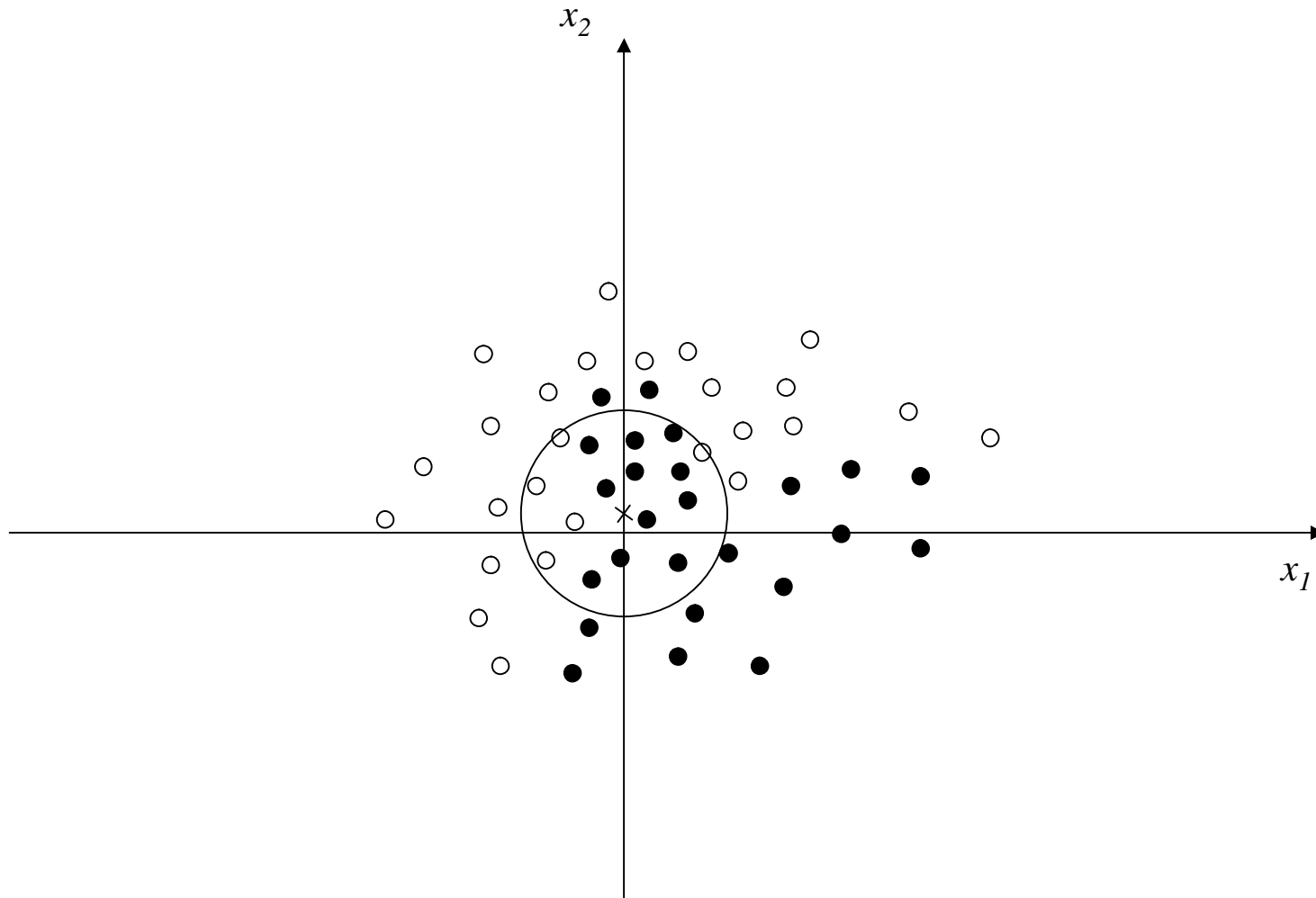
Supervised vs unsupervised analysis



Support vector machines

- The original n -dimensional space is transformed into even higher dimensional space using so-called *kernel function*
- The linear separation is done in that higher dimensional space and it is equivalent to a nonlinear separation in the original space

K nearest neighbours classifier



200069 .at; Q15020;
218139 .at; Q9H0R1;
21552 .at; Q9BV23; E...
45288 .at; Q9BV23; EN...
21679 .at; Q9BV23;
22710 .at; AAP88855;
21786 .at; Q724R8; E...
22754 .at; Q96Q11; E...
23584 .at; Q15006; E...
18889 .at; Q8WTT2; E...
27385 .at; Q8LY26;
22843 .at; Q86V18; E...
1558943 .at; AAH01610;
23893 .at; Q16394; E...
17842 .at; AAH50708;
29146 .at; Q15800; E...
24648 .at; Q86WP2; E...
25734 .at; AAH32540;
18254 .at; Q9H029;
19037 .at; Q9V389; E...
19913 .at; Q81XG0;
21298 .at; Q9H3T5;
14440 .at; P18440; E...
27916 .at; Q9M0T5;
15602 .at; AAH62363;
21396 .at; AAH65930;
21005 .at; ?
1565752 .at; AAH62363;
1565754 .at; AAH653;
235940 .at; Q8N2B1; E...
184348 .at; Q86GR2;
29798 .at; Q13632; E...
20992 .at; Q727T5;
26458 .at; Q94940;
2408 .at; Q94940;
25109 .at; Q96BK9; E...
2495 .at; CRE45819;
01111 .at; P55060; E...
25082 .at; CRE45991;
25706 .at; P11172;
15160 .at; P11172;
29983 .at; ?
23406 .at; AAH23557;
28869 .at; ?
22763 .at; AAH6348;
25881 .at; Q8N9T7;
26481 .at; Q8TBD9;
29981 .at; Q81WH4; E...
28737 .at; Q8N8N2; E...
29981 .at; P60983;
29588 .at; AAQ8929;
22666 .at; Q96919;
25580 .at; Q8N0T7;
29600 .at; Q15067;
21649 .at; Q8N0T5;
10145 .at; P47712; E...
28113 .at; AAH64133;
18722 .at; Q96ER9;
13186 .at; AAH56574;
18089 .at; Q96RV2; E...
24222 .at; P43489;
218799 .at; AAQ89223;
25510 .at; Q9HD40;
18239 .at; Q8N4P8;
20178 .at; Q9V311; E...
24093 .at; AAH0258;
26000 .at; ?
26000 .at; Q9NVR0; E...
23094 .at; Q15013; E...
18224 .at; Q9H0D4;
22623 .at; Q9UID6;
24777 .at; AAH64593;
33224 .at; ?
236484 .at; AAH14076;
1565689 .at; ?
23320 .at; Q96BMD; E...
23420 .at; Q9DBU2; E...
34993 .at; Q8NBW1; E...
24370 .at; Q92369; E...
19016 .at; ENSG000;
23657 .at; AAH62368;
222045 .at; BAC4523;
27864 .at; BAC5196;
209510 .at; AAH64636;
184036 .at; BAC23713;
154037 .at; BAC23713;
24780 .at; AAQ09533;
15563 .at; Q9H761; E...
24624 .at; Q9UID6;
24711 .at; AAH65365;
42343 .at; AAH63686;
24377 .at; Q8TBD9;
21235 .at; Q60466;
25376 .at; Q9NWU5; E...
208776 .at; Q00231; E...
1569320 .at; Q9H751;
35556 .at; ?
35556 .at; ?
13758 .at; AAH62437;
2735 .at; ?
27448 .at; Q9NWB6; E...
253015 .at; Q72530; E...
159121 .at; Q95376;
28480 .at; Q9POL0; E...
12241 .at; AAH62284;
12374 .at; Q9UK73; E...
25311 .at; AAH6489;
26650 .at; Q8NGM9; E...
22044 .at; BAC45238;
244648 .at; ?
159527 .at; ?
89948 .at; BAC45238; ...
43649 .at; Q9V311;
30134 .at; Q863T6;
29757 .at; ?
42688 .at; ?
42113 .at; ?
27373 .at; ?
28181 .at; Q96KD7; E...
42920 .at; ?
16094 .at; ?
24778 .at; ?
33360 .at; AAP35578;
36839 .at; ?
155436 .at; AAH630...

