References

1. Pattern Classification

Duda, Hart, Stork

Let $X \in \mathbb{R}^d$ with p.d.f. $f_X(x)$

$E(X) = \mu \qquad C_{ij} = E(X_i - \mu_i)(X_j - \mu_j)$

For any $u \in \mathbb{R}^d$

$z = u^T X$

Find $E(z)$, $var(z)$

$E(z) = E(u^T X) = u^T \mu$

$var(z) = E(u^T X - E(z))^2 = u^T C u$

Suppose $\|w\| = 1 \qquad \|w\| = \sqrt{w^T w}$

$C = \{\alpha w \mid \alpha \in \mathbb{R}\} \qquad = \sqrt{\sum_i w_i^2}$

Projection of any $v \in \mathbb{R}^d$ on C is the point in C which is closest to v

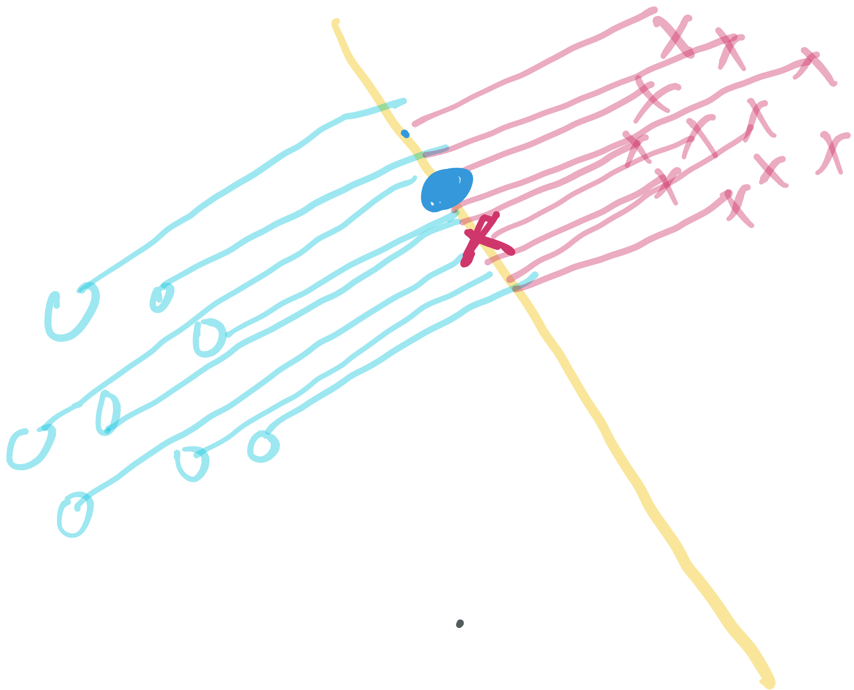
For any $v, w \in \mathbb{R}^d$

$\|w\|^2 \|v\|^2 \geq (v^T w)^2$

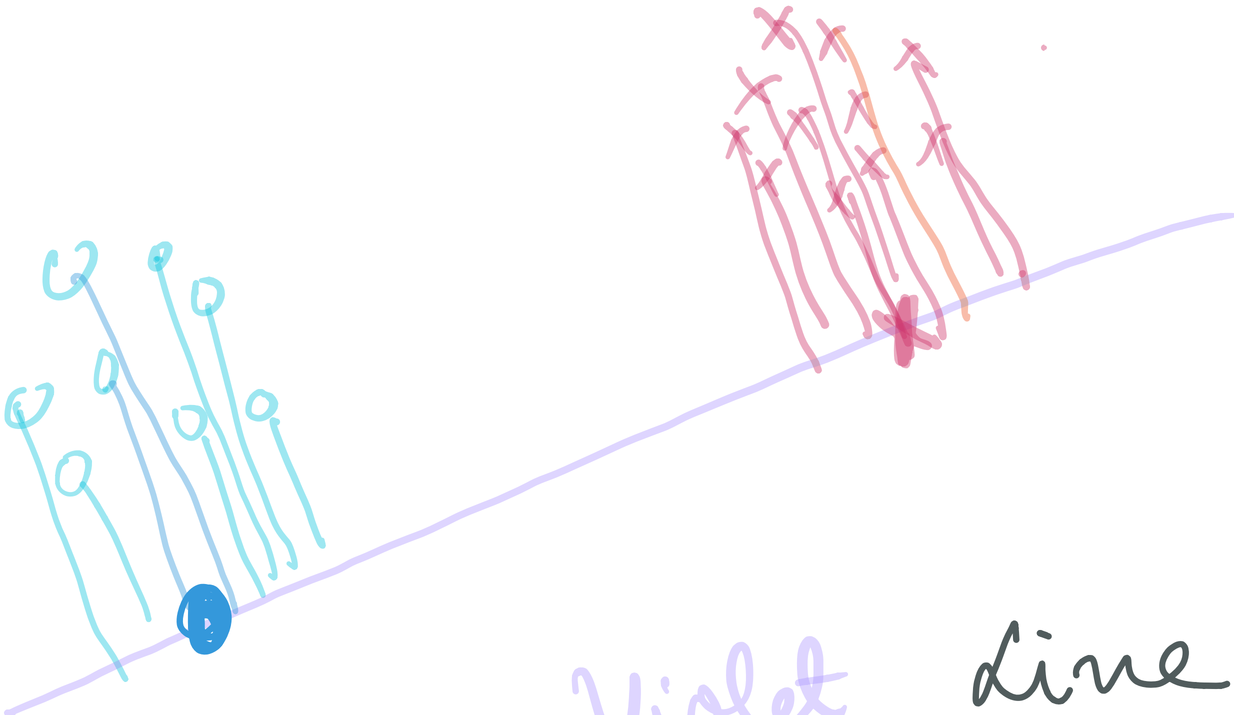
Equality holds iff $w = tv$

For any $x \in \mathbb{R}^d$

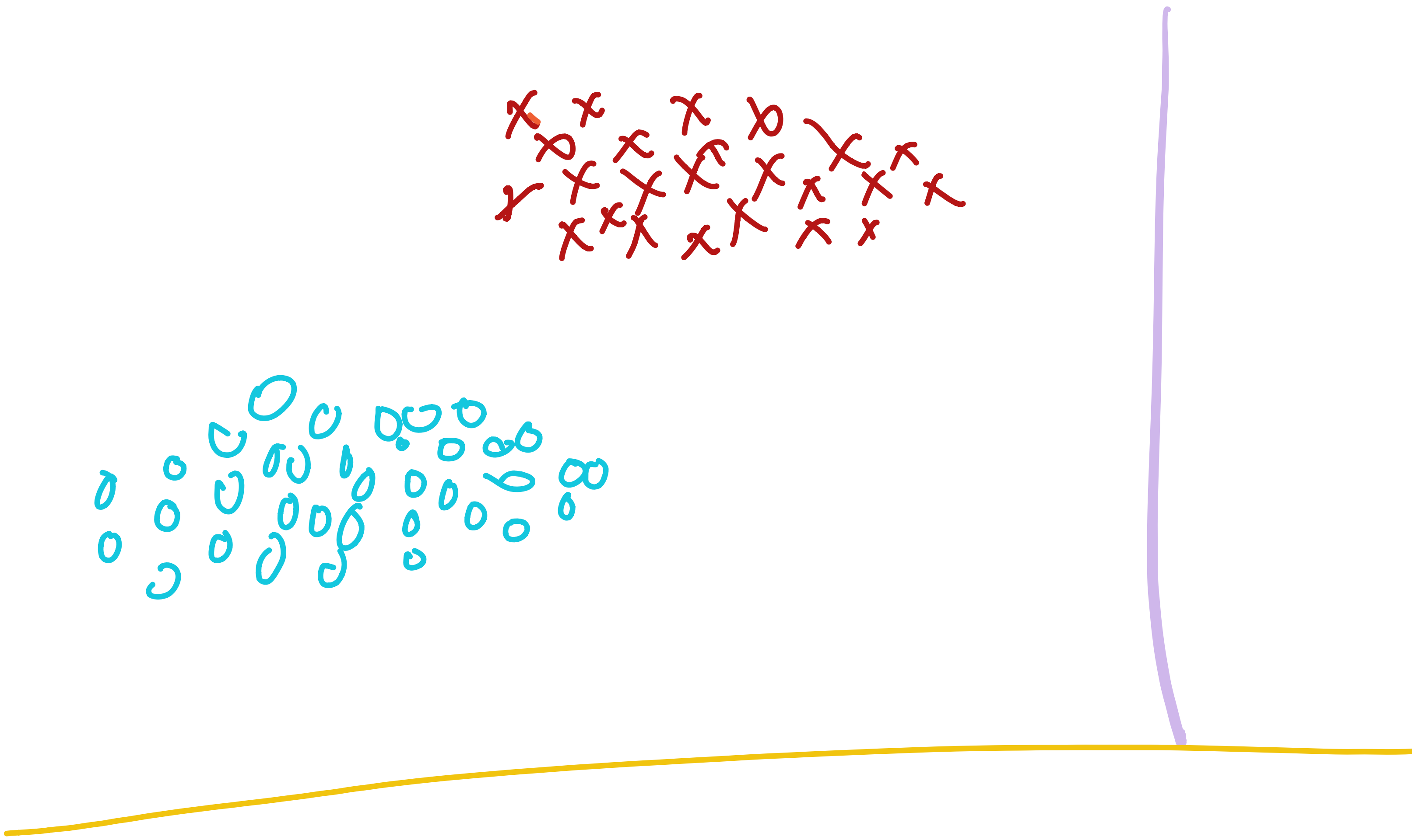
$w^T x$ is the projection of x on w, $(\|w\| = 1)$

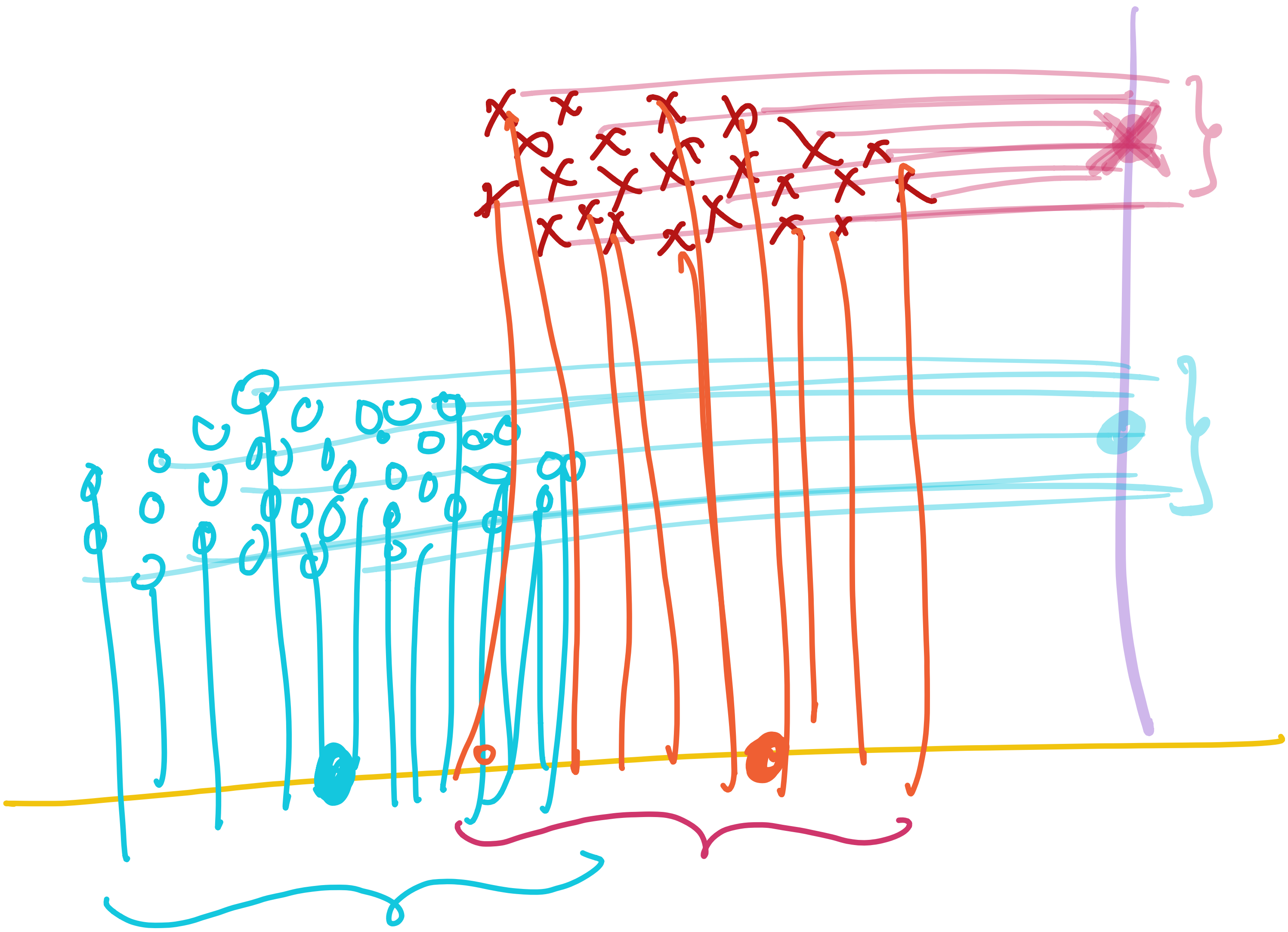


Overlap is Large

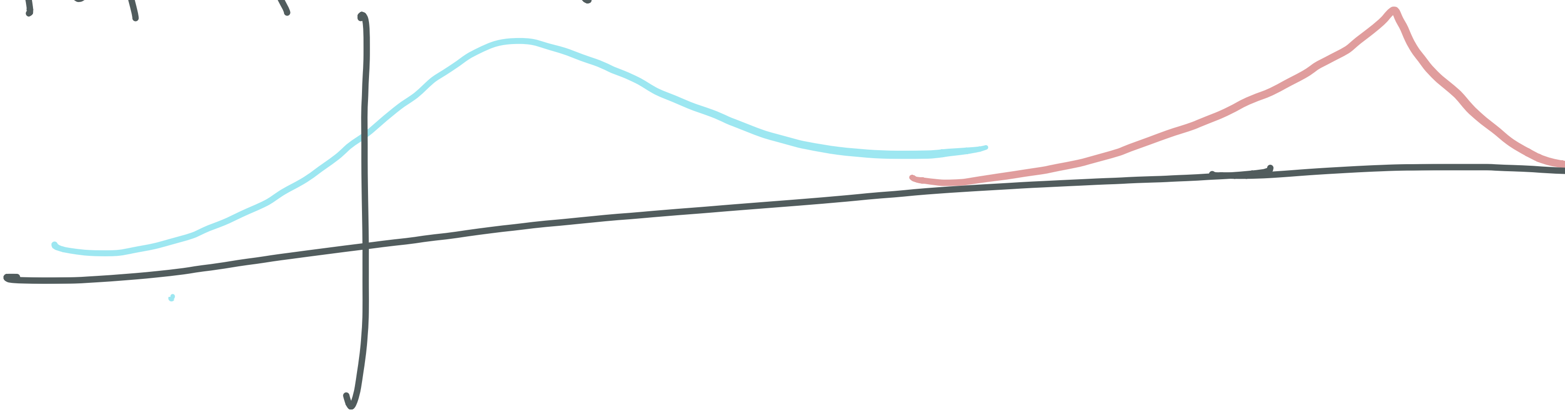


Projecting on Violet line
is much better than
Yellow line.



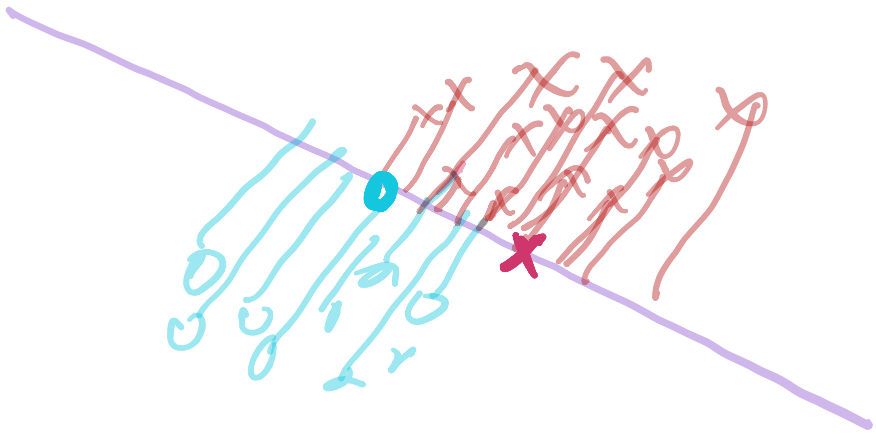


pdf of proj - ons

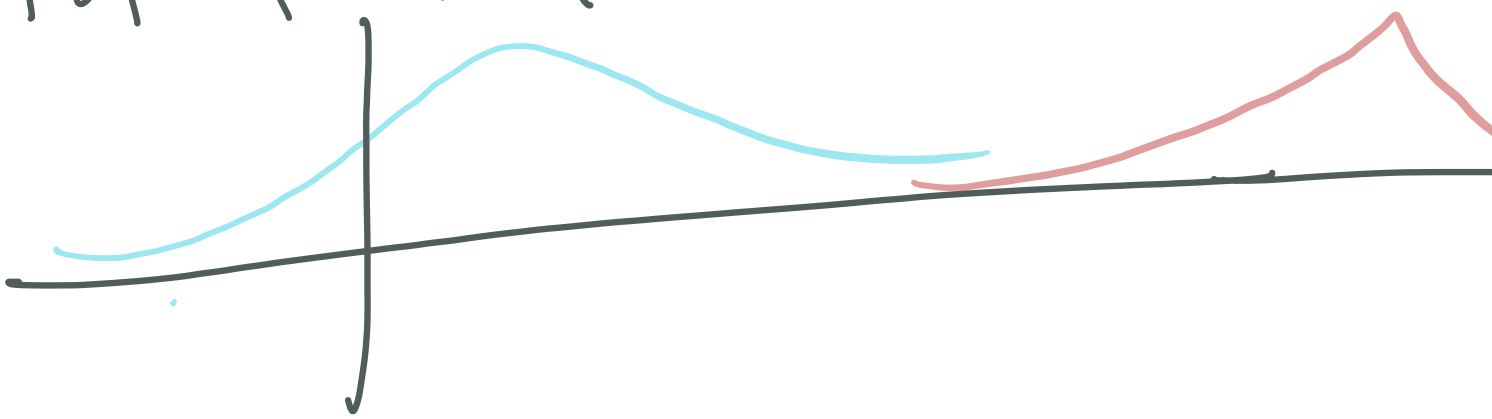


After projections mean should be well separated

variance should be small



pdf of proj - ons



After projections
should be well separated

$E(z) = w^T \mu$ $\quad$ $var(z) = w^T C w$

$|w^T \mu_1 - w^T \mu_2| \to$ Large

$w^T C_1 w$ and $w^T C_2 w$ small

$$\max_{w} \frac{|w^T(\mu_1 - \mu_2)|^2}{w^T C_1 w + w^T C_2 w}$$

No need for $\|w\| = 1$

$S_d^+ \to$ Symmetric, positive definite matrices

$A, B \in S_d^+$

$$\max_{w \in \mathbb{R}^d} \frac{w^T A w}{w^T B w}$$

$B = L^2 \qquad L \in S_d^+$

$L \Rightarrow B^{1/2}$

$u = B^{1/2} w. \qquad w = B^{-1/2} u$

$$\max_{u \in \mathbb{R}^d} \frac{u^T B^{-1/2} A B^{-1/2} u}{u^T u}$$

$$\max_{u \in \mathbb{R}^d} \frac{u^T E u}{u^T u} \quad E \in S_d$$

$$\max_{u \in \mathbb{R}^d} \frac{u^T E u}{u^T u} = \lambda_{max}(E) = \frac{u_0^T E u_0}{u_0^T u_0}$$

u_0 is the eigenvector of $\lambda_{max}(E)$

λ_{max} is the largest eigenvalue

$$\max_{u \in \mathbb{R}^d} \frac{u^T B^{-1/2} A B^{-1/2} u}{u^T u}$$

$$= \lambda_{max}\left(B^{-1/2} A B^{-1/2}\right) = \bar{\lambda}$$

attained at

$$B^{-1/2} A B^{-1/2} u_0 = \bar{\lambda} u_0$$

$$w = B^{-1/2} u$$

$$w_0 = \arg\max_{w \in \mathbb{R}^d} \frac{w^T A w}{w^T B w}$$

$$w_0 = B^{-1/2} u_0$$

$$\therefore \quad B^{-1} A w_0 = \bar{\lambda} w_0$$

$$\boxed{A w_0 = \bar{\lambda} B w_0}$$

Generalized eigenvalue problem

For Fisher discriminant

$$A = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

$$B = C_1 + C$$

$$(C_1 + C_2)^{-1} (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w_0$$
$$= \lambda w_0$$

$$(\mu_1 - \mu_2)^T (C_1 + C_2)^{-1} (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w_0$$
$$= \lambda (\mu_1 - \mu_2)^T w_0$$

Clearly then

$$\lambda = (\mu_1 - \mu_2)^T (C_1 + C_2)^{-1} (\mu_1 - \mu_2)$$

$$w_0 = (C_1 + C_2)^{-1} (\mu_1 - \mu_2)$$