Final exam info

- In-class exam
- Closed books/notes, closed electronics, 2 sided A4 hand-written (not printed) cheat sheet allowed (upload somewhere by some date)
- Videos on (contact instructor if there is issue by TOMORROW)
- Academic integrity violations will have severe consequences

- Multiple choice and Short answer
- Via Canvas
- Only 1 attempt allowed
- ~20 questions – 75 mins (3-4 mins/question)
  - ~40% are hard (think open book), ~60% are easy
- Grading will be curved
- Ask questions via private Zoom chat

$$\min_{w} p(w) \rightarrow L(w, \alpha)$$
$$\text{s.t. } g(w) \geq 0 \qquad \downarrow$$
$$\max_{\alpha} d(\alpha)$$

duality, KKT conditions

Topics (all, more emphasis since midterm)

- Basics - Probability, Matrix/vector calculus, Optimization (convexity etc)
- Basic ML concepts – training vs test data, overfitting, generalization, ML tasks, loss metrics, optimal classifier/regressor, decision boundaries, **model selection (hold out, cross-validation, model complexity), theory (Haussler, Hoeffding, VC dimension)**

- Unsupervised

  mixture models
  - Distribution/Density estimation – MLE, MAP, Histogram, Kernel density estimation
  - **Dimensionality reduction – PCA** $V \leftarrow \text{evec}(\underset{D\times n}{X}\underset{n\times D}{X^T})$ data centered
  - **Clustering – K-means, Gaussian mixture models (GMM), EM algorithm, Hierarchical**
- Supervised

  - Classification – Naïve Bayes, Logistic Regression, Neural Networks, k-Nearest Neighbors, **SVM, Kernel trick, Decision Trees, Boosting** $H(x) = \text{sign}\left(\sum_t \alpha_t h_t(x)\right)$
  - Regression – Linear Regression, Ridge, Lasso, Neural Networks, Kernel regression

    Kernelized ridge regression local

Comparison chart (classification)

Algorithm	Generative/ Discriminative	Assumptions	Decision boundary	Loss function	Training
Naïve Bayes					
Logistic Regression					
Neural Networks					
k-Nearest Neighbors					
SVM	D	-	-		Quad prog.
Decision Tree	D		-		inf gain ← ID3, C4.5 dynamic prog. ← CART (MDL, Inf(m))
Boosting	?				coordinate descent α_t, h_t on exp loss

Comparison (regression)

Algorithm	Generative/ Discriminative	Assumptions	Decision boundary	Loss function	Training
Linear Regression					
Neural Networks					
Local Kernel regression					
Kernel Ridge Regression					

Unsupervised learning

PCA

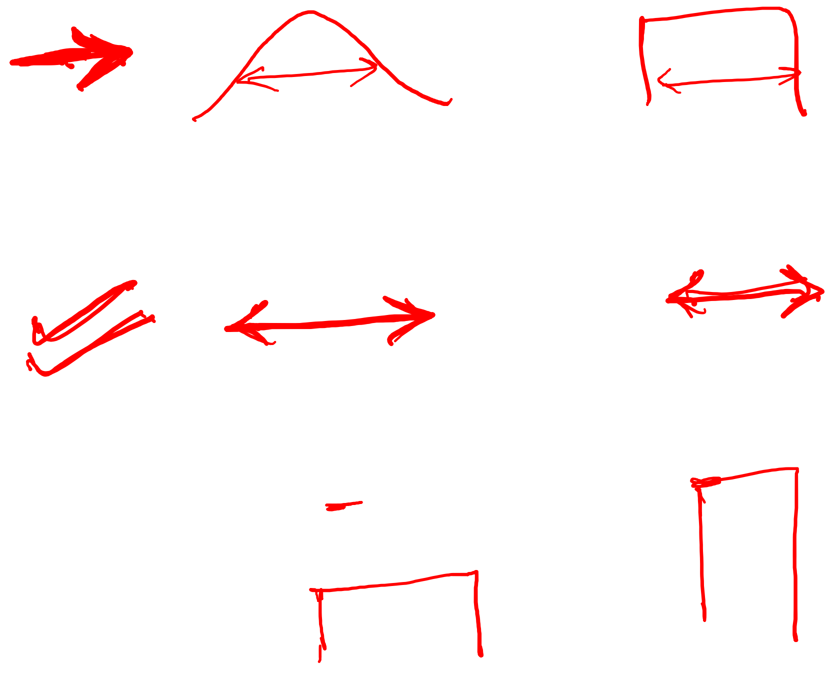
Clustering
- Partition (K-means, GMM+EM)
- Hierarchical

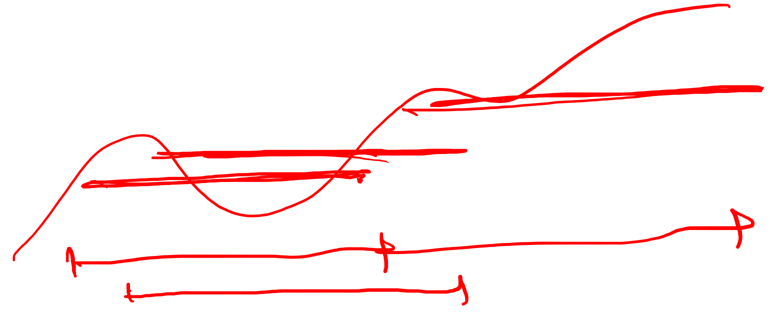
Distribution est.

Practice problems

$$K\left(\frac{\|x - x_i\|}{h}\right)$$

9.





$\|\cdot\|_2$ $\|\cdot\|_\infty$ $\|\cdot\|_1$

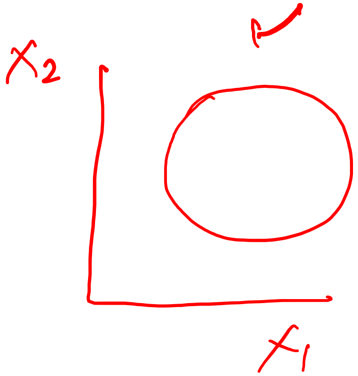
10.

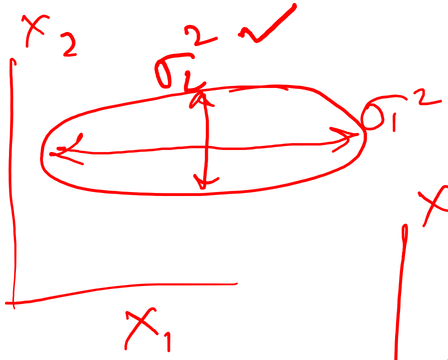
$$p(x \mid Y=y) \sim N(\mu_y, \Sigma_y) \qquad \Sigma_y = \begin{bmatrix} \cdot & \cdots & \cdot \\ \vdots & & \vdots \\ \cdot & \cdots & \cdot \end{bmatrix} \quad O(d^2)$$

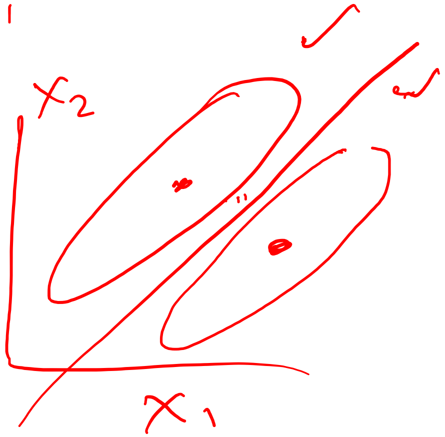
$$x = [x_1 - x_d]^T$$

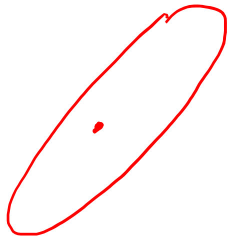
vs.

$$P(x \mid Y=y) = \prod_{j=1}^{d} p(x_j \mid Y=y) \Rightarrow \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \end{bmatrix} \quad O(d)$$









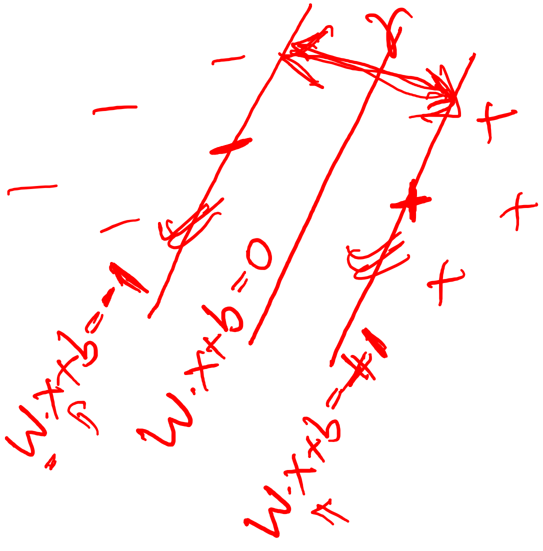
$2d$

d

$$\Phi_1(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$$

$$\Phi_2(x) = \begin{bmatrix} 2x \\ 2x^2 \end{bmatrix} \checkmark$$

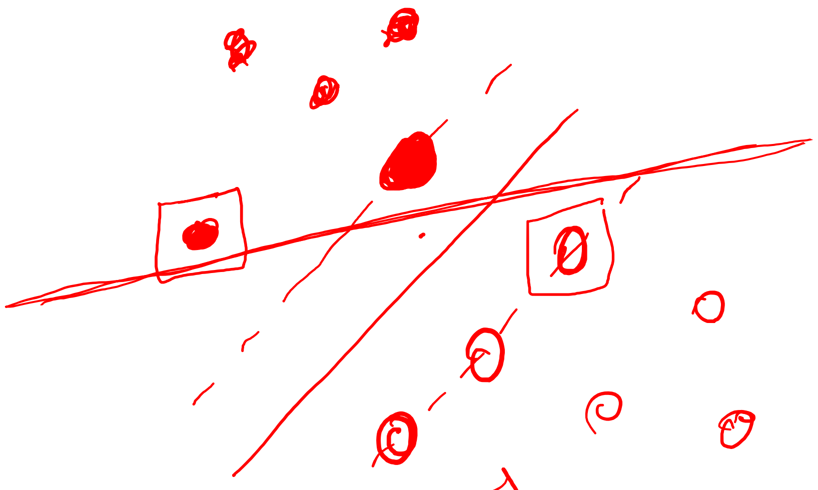
margin

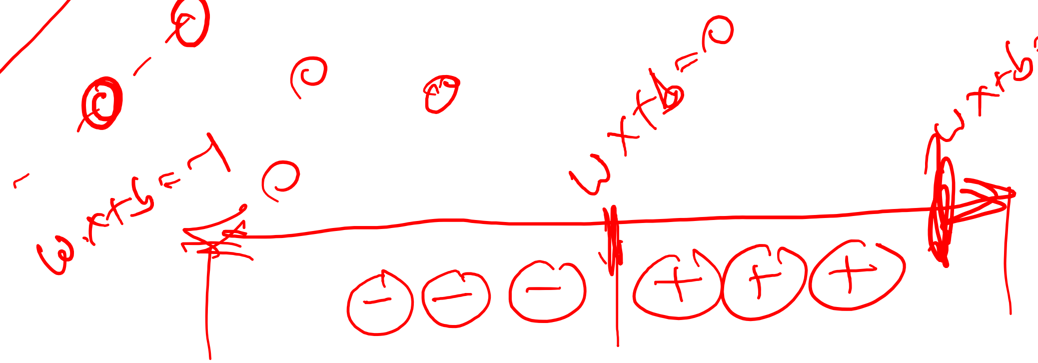


$$\gamma \propto \frac{1}{\|w\|^2}$$

$$\min \|w\|^2$$

17.





18.

$$\|w\|^2 + C \sum_j \xi_j \qquad C \to \infty$$

$$(w\cdot x_j+b)\, y_j \geq 1-\xi_j$$

$$\|w\|^2 + 0 \cdot \sum_j \xi_j$$

$$(w\cdot x_j+b)\, y_j \geq 1-\xi_j$$

$$\xi_j \geq 0$$

12.