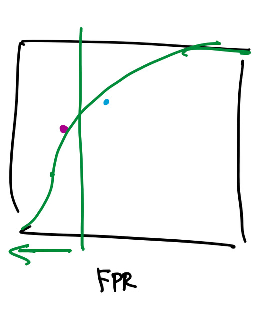
Lecture 24

Thursday, December 5, 2024 10:01 AM

Precision, recall, f1

[sensitivity (TPR), specificity (TNR), 1 − TNR = FPR, ROC]



logistic regression

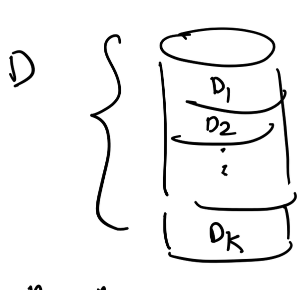
binary

$$P(y=P) = \sigma(\vec{\omega}^T \vec{x}) = \frac{1}{1+e^{-\vec{\omega}^T\vec{x}}} \in [0,1]$$

$$\hat{y} = \begin{cases} P & \text{if } P(y=P) \geq 0.5 \\ N & \text{otherwise} \end{cases}$$

---

k-fold cross validation



$n - \frac{n}{k}$

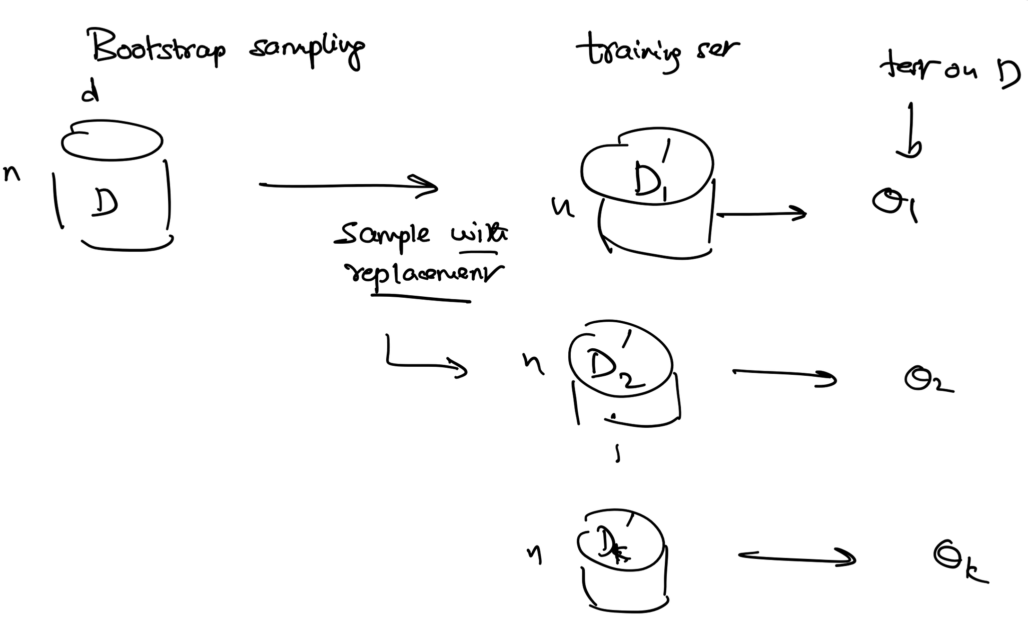
Use D_i as testing & $D - D_i$ as training → M_i (model for) → θ_i ← some eval metric (f1)

metrics from each of the folds → $\theta_1, \theta_2, \ldots, \theta_k$

$$\hat{\mu} = \frac{1}{k}\sum \theta_i$$

$$\hat{\sigma}^2 = \frac{1}{k}\sum (\theta_i - \mu)^2$$

Bootstrap sampling



from D: each point has prob $\frac{1}{n}$ of being selected for D' — μ, σ^2: optimistic

Prob of point not being selected is $\left(1 - \frac{1}{n}\right)$

Q. what's the prob that a given point is not selected even after n trials

$$\left(1 - \frac{1}{n}\right)^n \approx \frac{1}{e} = 0.368$$

⇒ Over between D & D' is 0.632 !

---

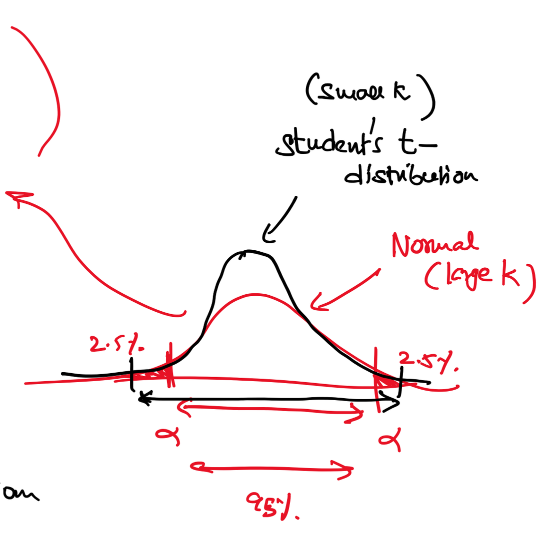
true mean performance — μ

predicted mean performance — $\hat{\mu} = \frac{1}{k}\sum \theta_i$

" var — $\hat{\sigma}^2 = \frac{1}{k}\sum(\theta_i - \mu)^2$

$$\left(\hat{\mu} - t_\alpha \frac{\hat{\sigma}}{\sqrt{k}} \leq \mu \leq \hat{\mu} + t_\alpha \frac{\hat{\sigma}}{\sqrt{k}} \right)$$

α is the confidence level → relates to the critical value of the Normal distribution



t-distribution with k−1 degrees of freedom

Comparing two models

	A	B
F1:	0.695	0.69
	0.595	0.60
	⋮	⋮

} k fold CV → $D - D_i$ | D_i → θ_i^A, θ_i^B → δ_i

$$\hat{\mu}_A = \frac{1}{k}\sum \theta_i^A, \quad \hat{\sigma}_A^2$$

$$\hat{\mu}_B = \frac{1}{k}\sum \theta_i^B, \quad \hat{\sigma}_B^2$$

Observe

$$\hat{\mu}_\delta = \frac{1}{k}\sum \delta_i, \quad \hat{\sigma}_\delta^2$$

$\delta_i = \theta_i^A - \theta_i^B$ — difference in performance on the same fold (paired t-test)

Hypothesis testing

Null hypothesis: there is no difference in the performance

truth

H_0: $\mu_\delta = 0$ (mean difference is 0)

H_a: $\mu_\delta \neq 0$ (A & B are statistically significantly difference)

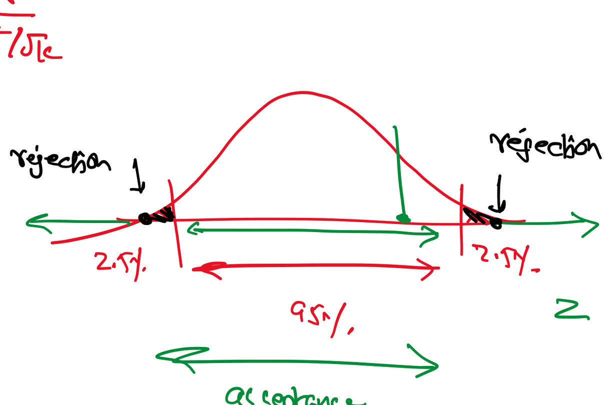
Under H_0:

$$Z_k = \frac{\hat{\mu}_\delta - \mu_\delta}{\hat{\sigma}/\sqrt{k}} = \frac{\hat{\mu}_\delta}{\hat{\sigma}/\sqrt{k}}$$

There is a significant difference if Z_k lies outside the α% interval

$z \notin [-t_\alpha, t_\alpha]$

the H_a is "true"



acceptance region

95% Z score is expected to lie in this region H_0

Exam II

No laptops !

↪ bring a calculator

Open nots / open book

Bayes classifier ⟺ KNN classifier

Discriminant Analysis

Linear regression

↪ via projections (new orthogonal basis)

$u_0 = \vec{1}$

$u_1 = X_1 -$ proj on prev

$u_2 = X_2 -$ proj on prev 2

no normalization

$\hat{Y} =$ Proj of Y onto u_0, u_1, u_2 — add them up

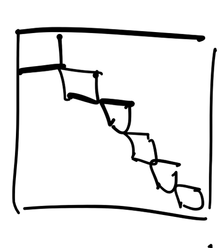
$\vec{\omega}$?

$$\hat{Y} = (D^T D)^{-1} D^T Y$$

+ ridge

Objective: $\sum \|y_i - \hat{y}_i\|^2 + \alpha \|\vec{\omega}\|_2$ (ridge)

$$\hat{Y} = (D^T D + \alpha I)^{-1} D^T Y$$



+ α

Logistic regression

$P(y=1) = \pi_1 = \sigma(\vec{\omega}^T \vec{x})$ ← sigmoid

odds ratio

log odd ratio $= \log\left(\frac{P(y=1)}{P(y=0)}\right)$

Maximum likelihood

$-\log \mathcal{L} = P(\vec{\omega} \mid D)$ → binary, k-way

$\pi_i = \text{softmax}\left(\vec{\omega}_i^T \vec{x} \mid \omega_1^T x, \omega_2^T x, \ldots \omega_k^T x\right)$

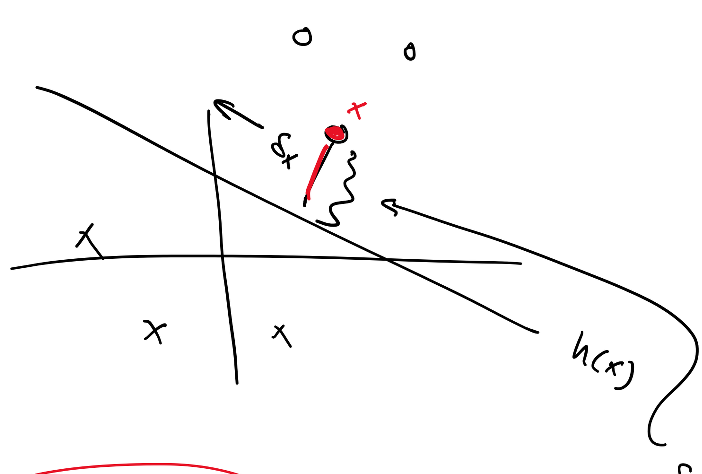
SVM

$\frac{1}{\|\omega\|}$ margin

Canonical hyperplane

$y = 2x + b$

$10y = 20x + 10b$

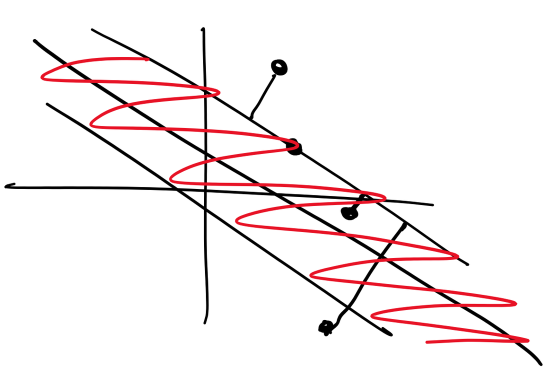


$$\delta_x = \frac{y \cdot h(x)}{\|\omega\|} = \frac{10 \cdot 2}{\|\omega\|}$$

$$\frac{y \cdot h(x)}{10 \cdot 2} = 1$$

$$w = \sum \alpha_i y_i \vec{x}_i$$

$$\frac{1}{2}\|\vec{\omega}\|^2 + C\sum \xi_i^k$$

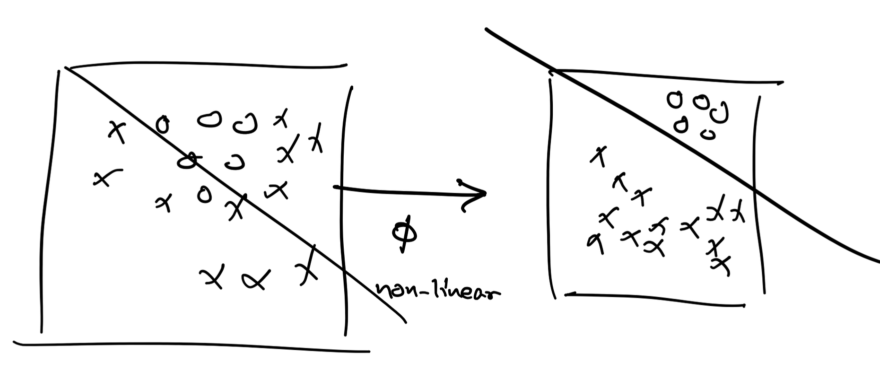


$\vec{\omega} = \sum_{\alpha_i > 0} \alpha_i y_i \vec{x}_i$ — support vectors

$\alpha_i \geq 0$

$$\vec{\omega} = \sum_{\alpha_i > 0} \alpha_i y_i \phi(\vec{x}_i)$$

transformed point



ϕ → gaussian function

$\vec{x} \to \phi(\vec{x}, \cdot)$

$\mathbb{R}^d \to \mathbb{R}^\infty$

$$\text{Classifier} = \text{sign}(\phi(z)^T \omega)$$

$$= \text{sign}\left(\phi(z)^T \left(\sum \alpha_i y_i \phi(x_i)\right)\right)$$

$$= \text{sign}\left(\sum \alpha_i y_i \phi(z)^T \phi(x_i)\right)$$

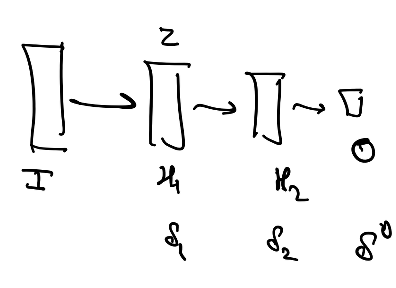
$$= \text{sign}\left(\sum_{i=1}^n \alpha_i y_i k(z, x_i)\right)$$

$$k(z, x_i) = e^{-\frac{\|\vec{z} - \vec{x}_i\|^2}{2\sigma^2}}$$

$$k(z, x_i) = (c + z^T x_i)^q$$

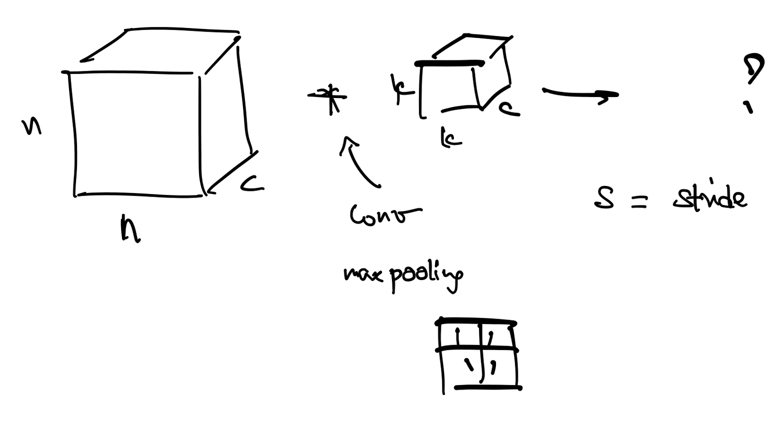
MLP

net gradients $\delta = \frac{\partial \mathcal{E}}{\partial net}$



$\nabla_\omega = z^{l-1} \cdot \delta^T$, $\nabla_b = \delta$

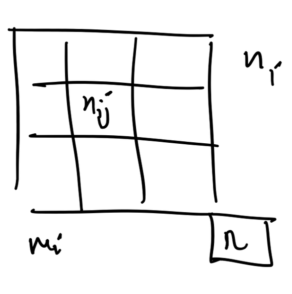
CNN



s = stride

Assessment

Confusion matrix



roc - curve (5 points / score)

confidence interval: t_α value for different α, k−1 dof

$\hat{\mu}$, $\hat{\sigma}$

(μ ?)

Q. is there a significant difference