ML Week 6 - Model Valid 2

last week = wider picture of data problem
preprossing, Collecta,

today = More preprocessing & Evalulatia

- Dimensionality Reduction
- Evaluation metrics
- Evaluatia strategies

Recap Eigenvalues & Eigenvectors

Eigens stretch or squish an arrow vector

for any square matrix $A \in \mathbb{R}^{B \times B}$
there exists a non-zero vector $v \in R^B$ such that:

$$Av = \lambda v$$

λ = Scalar

the scalar has the same effect as the matrix transformatia

v is the eignvector of A
& λ is the eigen value

Eigen vectors & values are useful for analysij a matrix in the context of linear mapping

One of their uses is Eigendecomposition

A square matrix $A \in \mathbb{R}^{B \times B}$ can be decomposed as

$$A = P \cdot D \cdot P^{-1}$$

P = Matrix of eigenvectors of A $\begin{bmatrix} v_1^1 & \cdots & v_1^B \\ \vdots & \ddots & \vdots \\ v_B^1 & \cdots & v_B^B \end{bmatrix}$

D = A Diagonal matrix whose entries are the corresponding eigenvalues of A $\begin{bmatrix} \lambda^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda^B \end{bmatrix}$

this only true if ~~the~~ eigenvectors are lin independant

that is, none can be written as a linear combination of the others

Dimensionality Reduction

Input w/ 10k Dims

memory & time concerns
(Complexity)

Model complexity = concern about overfitting

Solution
→ Reduce dims of train data
→ But in a way to not loose ~~va~~ info valuable for the MC task

Approaches:

* Based on Domain knowledge (Hand Craft features)
  - Feature engineering

* Exhaustive comparison of DIFF combs of features
  - Not efficient & often possible for large dims
  - Could be useful for Dims that have already been pruned using Domain knowledge

* Representation / feature learning (week 9)

* Other methods
  - PCA ⟶ today lecture

    ↳ does not reduce itself
    → Maps data into diff spaces

Principal Component Analysis (PCA)

maps input data $x \in \mathbb{R}^D$ to $\tilde{x} \in \mathbb{R}^D$

such that the new $\tilde{x}$ (PC's) are ordered

1st PC is most informative
- captures largest variance in orig data

2nd PC next most informativ
- captures largest var not captured by 1st PC

3rd etc...

Dim Reduction goal is to:
* Make data dims more compact (reduced)
* w/ minimal loss of information (variance)

reframe this as:
transform x into new space where a single feature $\tilde{x}_1$ captures the max variance in the original space

$$\tilde{x}_{1,n} = b_1^T x_n \qquad \forall n, n = 1, 2, \ldots, N$$

and maximising
$$V_{\tilde{x}_1} = \frac{1}{N} \sum_{n=1}^{} \tilde{x}_{1,n}^2 \qquad \text{where } V_{\tilde{x}_1} = \text{var of } \tilde{x}_1$$

$b_1 = \in B$ set of orthonormal basis vectors that map x_n to $\tilde{x}_n$

Orthonormal means orthogonal vectors, each normalized to unit length

Stopped taking notes of Dimensionality Reduction as readings will be better

Next Section:

Evaluation metrics

Accuracy → Classification, proportion of correct classifications

$$= \sum_{n=1}^{N} \begin{cases} 1 & \text{if } \hat{y}_n = y_n \\ 0 & \text{otherwise} \end{cases} \Big/ N \quad = \quad 0 \leq acc \leq 1$$

0 = total wrong, 1 = total non

How to judge accuracy?
* Random guess = lower Baseline → $\frac{1}{N}$
* Compare to human performance / Prediction

lots of metrics in weeks 3/4 labs

Confusion Matrix

A repres of classif performance

		Pred 0	Pred 1
True 0		TN	FP
True 1		FN	TP

$$ACC = \frac{TP + TN}{TN + FP + FN + TP}$$

Can be extened to Multiclass

Problem w/ accuracy

- Report accuracy but never on its own
- is a biased metric → hides errors & deficiencies
- confusion matrix are most transparent

↳ but sometimes you need a single metric

Options → Precision, recall, f1 score

Precision: Prop of predicted class that are correct
$= TP / (TP + FP)$

Recall: Prop of true class that are correct
$= TP / (TP + FN)$

these two are conflicting metrics, but aim of to opt both

f1 Score captures the balance between them

$$f1 = 2 \times \frac{Prec \times Recall}{Prec + Recall} = \frac{2TP}{2TP + FP + FN}$$

Can/should compute f1 for each class

Other Optios

- Roc Curve = True pos rate vs false pos rate

AUC = Area under curve = single metric

Plots the different outcomes for diff Decision Boundry
→ is probabalistic boundries from model

Other metrics

- Correlation metrics
- Specialist metrics
- Common regression metrics

Next section:

Evaluation Strategies

Divide data into train + test ← for eval

How to split?

Hold out validation ← week 2
80/20

Problem? Bias splits

k-fold cross valid
- k = splits/fold
  - each time use one fold as test as iterate through folds

- less biases

$k = N$

- higher level of k-cross = leave one out valid
  → max size of train
  → slow + costly

k = 10 = typical

Alternatives to k-cross?

Split data by groups, i.e. people

i.e. leave out = all data for partic group

then can also take a k-cross appour for group

basically as hand selected k-cross

How to report k-cross valid?

opt 1 - · Report mean & SD of folds
· cant use confusion Matrix

opt 2 - • Combine all test labels & preds accross folds
• then, compute metric

Validation Set - week 3

train → test → valid

↓ ↓ ↓

train | Evaluation | optimize hyperparas Week 4
• early stopping epochs