

CSC 144V: Autonomous Vehicles

Lecture 2.1

Intro to Machine Learning Exercises Solution

Department of Computer Science,
Hofstra University

Exercise: ML taxonomy

Classify each task in an autonomous-driving stack: supervised classification (C), supervised regression (R), unsupervised (U) or reinforcement learning (RL). Then say whether the labelled data come from a "teacher". Finally: which step runs on the car, training or inference, and which needs a GPU cluster?

Task	Type	Why
Recognize a traffic sign (stop, yield, speed limit ...)		
Predict the steering angle from a camera image		
Group similar driving scenes without any labels		
Learn a lane-change policy from a reward in a simulator		
Predict where a pedestrian will be in 2 s (x, y)		

Exercise Solution: ML taxonomy

- **Training** (millions of iterations of forward pass + back-propagation, days to weeks on GPU clusters) is done offline; **inference** is a single forward pass and runs on the car (edge device).
- Supervised = labelled examples from a "teacher"; unsupervised = no labels, find structure; RL = no labelled answers, only a **reward** received after actions.

Task	Type	Why
Recognize a traffic sign (stop, yield, speed limit ...)	C	discrete labels
Predict the steering angle from a camera image	R	continuous value
Group similar driving scenes without any labels	U	no labels, find structure
Learn a lane-change policy from a reward in a simulator	RL	maximize cumulative reward
Predict where a pedestrian will be in 2 s (x, y)	R	continuous values

Exercise: Linear regression and loss

Fit a line $\hat{y} = w \cdot x + b$ by least squares to the points (1, 2), (2, 3), (3, 5), (4, 6). Use $w = \Sigma(x - \bar{x})(y - \bar{y}) / \Sigma(x - \bar{x})^2$ and $b = \bar{y} - w \cdot \bar{x}$.

- (a) Compute $\bar{x}$, $\bar{y}$, w , b . (b) Fill in the table and compute the MSE and the MAE. (c) Predict $\hat{y}$ at $x = 5$. (d) Two models both have the absolute errors (1, 1, 1, 1) and (0, 0, 0, 4). Compare MSE and MAE of the two: which loss punishes the single large error more?

x	y	$\hat{y} = 1.4x + 0.5$	residual $y - \hat{y}$	squared error
1	2			
2	3			
3	5			
4	6			

Exercise Solution: Linear regression and loss

- (a) $\bar{x} = 2.5$, $\bar{y} = 4$; $\Sigma(x - \bar{x})(y - \bar{y}) = 7$, $\Sigma(x - \bar{x})^2 = 5$, so **$w = 1.4$, $b = 0.5$** . (c) $\hat{y}(5) = 1.4 \cdot 5 + 0.5 = \mathbf{7.5}$.
- (b) $MSE = (0.01 + 0.09 + 0.09 + 0.01)/4 = \mathbf{0.05}$, $MAE = (0.1 + 0.3 + 0.3 + 0.1)/4 = \mathbf{0.2}$.
- (d) Errors (1,1,1,1): MAE 1, MSE 1. Errors (0,0,0,4): MAE 1, MSE **4**. MSE (L2 loss) squares the error and punishes outliers more; MAE (L1) treats all errors linearly and is more robust.

x	y	$\hat{y} = 1.4x + 0.5$	residual $y - \hat{y}$	squared error
1	2	1.9	0.1	0.01
2	3	3.3	-0.3	0.09
3	5	4.7	0.3	0.09
4	6	6.1	-0.1	0.01

Exercise: Logistic regression

Logistic regression for "obstacle ahead?": $p = \sigma(w \cdot x + b)$ with $w = (1.5, -2)$, $b = 0.5$ and $\sigma(s) = 1/(1 + e^{-s})$. Predict class 1 if $p > 0.5$. Three labelled samples are given.

(a) Fill in s , p , the predicted class and the binary cross-entropy loss $-[y \ln p + (1 - y) \ln(1 - p)]$. (b) Which samples are misclassified? (c) Where is the decision boundary?

#	x	y	$s = w \cdot x + b$	$p = \sigma(s)$	Pred.	Loss
1	(2, 1)	1				
2	(0, 1)	0				
3	(1, 0.75)	0				

Exercise Solution: Logistic regression

- (b) Sample 3 is **misclassified**: $p = 0.62 > 0.5$ but $y = 0$. It also has the largest loss: a wrong, fairly confident prediction costs much more than a correct one.
- (c) $p = 0.5$ exactly when $s = 0$, i.e. $1.5 \cdot x_1 - 2 \cdot x_2 + 0.5 = 0$, a straight line in the feature space: logistic regression is a **linear** classifier (σ only turns the score into a probability).

#	x	y	$s = w \cdot x + b$	$p = \sigma(s)$	Pred.	Loss
1	(2, 1)	1	1.5	0.818	1	0.201
2	(0, 1)	0	-1.5	0.182	0	0.201
3	(1, 0.75)	0	0.5	0.623	1	0.974

Exercise: Model capacity and overfitting

A polynomial of degree d is fit to the same 30 training samples, and the error is measured on 100 unseen test samples.

(a) Fill in the gap (test – train) and the diagnosis. (b) Which degree would you choose? (c) What does a big gap between training and validation accuracy suggest, and what does no gap (but a high error) suggest? What do you change in each case?

Degree d	Train MSE	Test MSE	Gap	Diagnosis
1	4.0	4.3		
3	0.9	1.1		
9	0.05	6.5		

Exercise Solution: Model capacity and overfitting

- (b) **d = 3**: lowest test error. $d = 9$ reaches almost zero training error but memorizes the noise.
- (c) **Big gap $\rightarrow$ overfitting** $\rightarrow$ decrease capacity (fewer layers/neurons, regularization, more data). **No gap but high error $\rightarrow$ underfitting** $\rightarrow$ increase capacity. More capacity always lowers the training error, but not the test error.

Degree d	Train MSE	Test MSE	Gap	Diagnosis
1	4.0	4.3	0.3	underfitting
3	0.9	1.1	0.2	good fit
9	0.05	6.5	6.45	overfitting

Exercise: Confusion matrix and metrics

A pedestrian detector is tested on 1,000 camera frames, 100 of which contain a pedestrian (positive class). It finds $TP = 80$, produces $FP = 45$ false alarms and misses $FN = 20$.

(a) Compute TN . (b) Compute accuracy, precision, recall and the F1 score. (c) A "detector" that always answers "no pedestrian": what is its accuracy and recall?

(d) For a pedestrian detector, which is worse, an FP or an FN ? How should the decision threshold move?

	Predicted positive	Predicted negative
Actual positive (100)	$TP = 80$	$FN = 20$
Actual negative (900)	$FP = 45$	

Exercise Solution: Confusion matrix and metrics

- (a) $TN = 1000 - 80 - 45 - 20 = 855$. (b) Accuracy = $935/1000 = 93.5 \%$, precision = $80/125 = 0.64$, recall = $80/100 = 0.80$, F1 = $2PR/(P + R) = 0.711$.
- (c) Accuracy **90 %** but recall **0**: accuracy is misleading with imbalanced classes, so use precision, recall and F1.
- (d) A missed pedestrian (**FN**) is usually worse than a false alarm (**FP**, e.g. unnecessary braking). Lowering the threshold raises recall (and TPR) but also raises FPR.

	Predicted positive	Predicted negative
Actual positive (100)	TP = 80	FN = 20
Actual negative (900)	FP = 45	TN = 855

Exercise: ROC curve and AUC

The same detector outputs a score; changing the decision threshold gives different operating points ($P = 100$ positives, $N = 900$ negatives).

(a) Fill in $TPR = TP/P$ and $FPR = FP/N$. (b) Estimate the AUC from the ROC points $(0, 0)$, the three points, and $(1, 1)$ with the trapezoid rule. (c) What does $AUC = 0.5$ mean, and what does a point at the top-left corner mean?

Threshold	TP	FP	TPR	FPR
0.9	50	9		
0.5	80	45		
0.1	95	270		

Exercise Solution: ROC curve and AUC

- (b) Trapezoids: $0.01 \cdot 0.25 + 0.04 \cdot 0.65 + 0.25 \cdot 0.875 + 0.70 \cdot 0.975 = 0.0025 + 0.026 + 0.219 + 0.683 \approx \mathbf{0.93}$.
- (c) AUC = 0.5: the score is no better than **random guessing** (the diagonal). Top-left (FPR = 0, TPR = 1) is the **perfect classifier**. AUC is independent of the threshold; ROC shows the TPR/FPR trade-off.

Threshold	TP	FP	TPR	FPR
0.9	50	9	0.50	0.01
0.5	80	45	0.80	0.05
0.1	95	270	0.95	0.30

Exercise: Validation and tuning

Choosing the learning rate with 5-fold cross-validation on the training set. The test set is kept aside.

(a) Compute the mean validation accuracy of each candidate and pick the best. (b) Why must the test set not be used for this choice? (c) With 9 trials and 2 hyperparameters, how many distinct values of each does a 3×3 grid search try, and how many does random search try? (Bergstra & Bengio, 2012)

Learning rate	Fold 1–5 accuracy (%)	Mean (%)
0.001	80, 82, 81, 79, 83	
0.01	86, 88, 85, 87, 89	
0.1	60, 55, 70, 58, 62	

Exercise Solution: Validation and tuning

- (b) Choosing on the test set would **leak** it into the model selection: the test accuracy would be optimistic. Use the validation folds to choose, and the test set **once** at the end.
- (c) Grid: **3** values per hyperparameter. Random: **9** distinct values of each, so important hyperparameters are explored better for the same budget.

Learning rate	Fold 1–5 accuracy (%)	Mean (%)
0.001	80, 82, 81, 79, 83	81
0.01	86, 88, 85, 87, 89	87 (best)
0.1	60, 55, 70, 58, 62	61