

Lecture 2 Linear Classification

Fei Tan

Department of Economics
Chaifetz School of Business
Saint Louis University

E6930 Introduction to Artificial Neural Networks

June 16, 2024

Image Classification



08	02	22	97	38	15	00	40	00	75	04	05	07	78	52	12	50	77	51	88
49	49	99	40	17	81	18	57	60	87	17	40	98	43	60	15	04	56	62	00
81	49	31	73	55	79	14	29	93	71	40	87	58	88	30	03	49	13	36	65
52	70	95	23	04	60	11	42	60	44	85	56	01	32	56	71	37	02	36	91
22	31	16	71	51	67	87	89	41	92	36	54	22	40	40	28	66	33	13	80
24	47	31	60	99	03	45	02	44	75	33	53	78	36	84	20	35	17	12	50
42	98	81	28	64	23	67	10	26	38	40	67	59	54	70	66	18	38	64	70
67	26	20	68	02	62	12	20	95	63	94	39	43	08	40	91	66	49	94	21
24	55	58	05	66	73	99	26	97	17	78	78	96	83	14	88	34	89	63	72
21	36	23	09	75	00	76	44	20	45	35	14	00	61	33	97	34	31	33	95
78	17	53	28	22	75	31	67	15	94	03	80	04	62	16	14	09	53	56	92
16	39	05	42	96	35	31	47	55	58	88	24	00	17	54	24	36	29	85	57
86	56	00	48	35	71	89	07	05	44	44	37	44	60	21	58	51	54	17	58
19	80	81	68	05	94	47	69	28	73	92	13	86	52	17	77	04	89	55	40
04	52	08	83	97	35	99	16	07	97	57	32	16	26	26	79	33	27	98	66
37	46	68	87	57	62	20	72	03	46	33	67	46	55	12	32	63	93	53	69
04	42	16	73	35	85	39	11	24	94	72	18	08	46	29	32	40	62	76	36
20	69	36	41	72	30	23	88	34	46	83	69	82	67	59	85	74	04	36	16
20	73	35	29	78	31	90	01	74	31	49	71	48	64	81	16	23	57	05	54
01	70	54	71	83	51	54	69	16	92	33	48	61	43	52	01	89	54	62	48

What the computer sees

→
image classification
82% cat
15% dog
2% hat
1% mug

The Road Ahead...

- ▶ Score function: parameterization, interpretation
- ▶ Loss function: data loss, regularization/prior
- ▶ SVM vs. Softmax classifiers

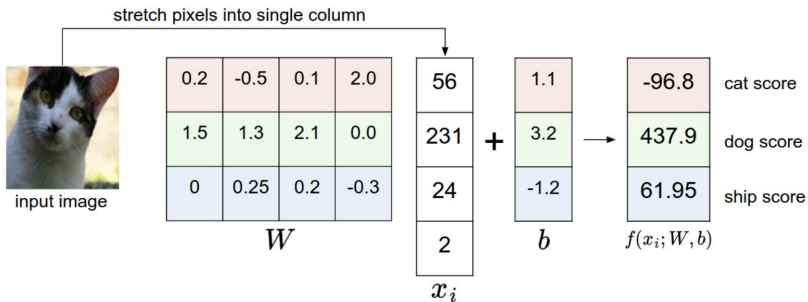
Score Function

Linear score

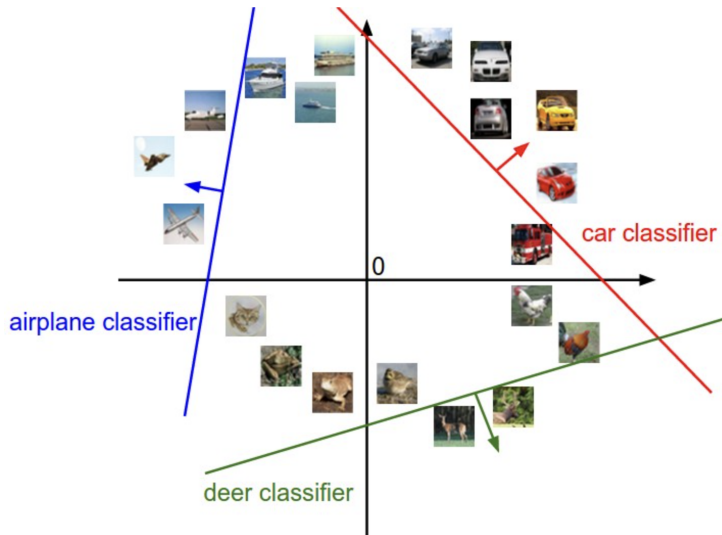
$$f(x_i, W, b) = Wx_i + b$$

- ▶ Notation
 - ▶ dataset of N examples $\{(x_i, y_i)\}_{i=1}^N$, D (normalized) features $x_i \in \mathbb{R}^D$, K categories $y_i \in 1, \dots, K$
 - ▶ $K \times D$ weights W , $K \times 1$ bias b
- ▶ Pipeline
 - ▶ split data into training, validation, and test sets
 - ▶ train parameters and (cross) validate hyperparameters
 - ▶ evaluate on test test only once at the end

Algebraic Interpretation



Geometric Interpretation



Loss Function

Regularized loss

$$L = \frac{1}{N} \sum_i L_i + \lambda R(W), \quad \lambda > 0$$

- ▶ Examples of data loss
 - ▶ multiclass support vector machine (SVM) classifier uses hinge loss: $L_i = \sum_{j \neq y_i} \max(0, f_j - f_{y_i} + \Delta)$, $\Delta > 0$
 - ▶ Softmax classifier uses cross-entropy loss: $L_i = -\log \left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}} \right)$
- ▶ Examples of regularization/prior
 - ▶ L^1 regularization: $R(W) = \sum_{i,j} |W_{ij}|$
 - ▶ L^2 regularization: $R(W) = \sum_{i,j} W_{ij}^2$

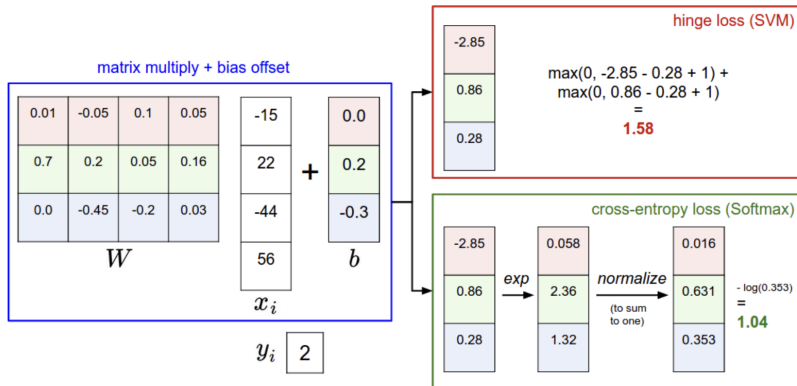
Loss Function (Cont'd)

```
import numpy as np

# SVM classifier
def L1_i(x_i, y_i, W, b):
    delta = 1.0
    f = W.dot(x_i) + b
    margins = np.maximum(0, f - f[y_i] + delta)
    margins[y_i] = 0 # ignore true class
    return np.sum(margins)

# Softmax classifier
def L2_i(x_i, y_i, W, b):
    f = W.dot(x_i) + b
    f -= np.max(f) # avoid potential blowup
    p = np.exp(f) / np.sum(np.exp(f))
    return -np.log(p[y_i])
```


SVM vs. Softmax



References

- ▶ cs231n.stanford.edu – CS231n: Deep Learning for Computer Vision
- ▶ Tang (2013), “Deep Learning using Linear Support Vector Machines”, *arXiv:1306.0239*