L1 distance:

$$d_1(I_1, I_2) = \sum_P |I_1^P - I_2^P|$$

k means : take closest k points vote.

L2: distance: $$d_2(I_1, I_2) = \sqrt{\sum_P (I_1^P - I_2^P)^2}$$

Split : train, validate, test data.

Linear classifier:

$$f(x, W) = Wx + b.$$

Multiclass SVM loss:

score vector: $s = f(x_i, W):$

$$L_i = \sum_{j \neq y_i} \begin{cases} 0 & \text{if } s_{y_i} \geq s_j + 1 \\ s_j - s_{y_i} + 1 & \text{otherwise} \end{cases} = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

$\underset{\text{safty margin}}{\uparrow}$

Regularization:

$$L(W) = \frac{1}{N} \sum_{i=1}^{N} L_i(f(x_i, W), y_i) + \lambda R(W)$$

L2: $R(W) = \sum_k \sum_l W_{k,l}^2$

L1: $R(W) = \sum_k \sum_l |W_{k,l}|$

Elastic net (L1 + L2): $\sum_k \sum_l |W_{k,l}| + \beta W_{k,l}^2$

Gradient descent.

omitted   sive have taken CS 22f