Interactive Interview Session

[top product based companies]

AppliedAICourse.com

→ sample of questions & variations.

→ 1 or 2 rounds of programming

4-5 rounds of DS/ML/DL/Math [depth of understanding]

↳ Depth, breadth, Foundational Math, Applied, Problem-solving etc.

→ submit your answers/ideas in the chat window

→ role-dependent [DA, DS, MLS, MLE, SDE-ML]

→ follow-up questions

→ behavioral (and) past-experience (and) projects based

→ open-ended questions : build Alexa's wake word detection [skipped]

→ skipped straight-forward questions (e.g explain GBDT)

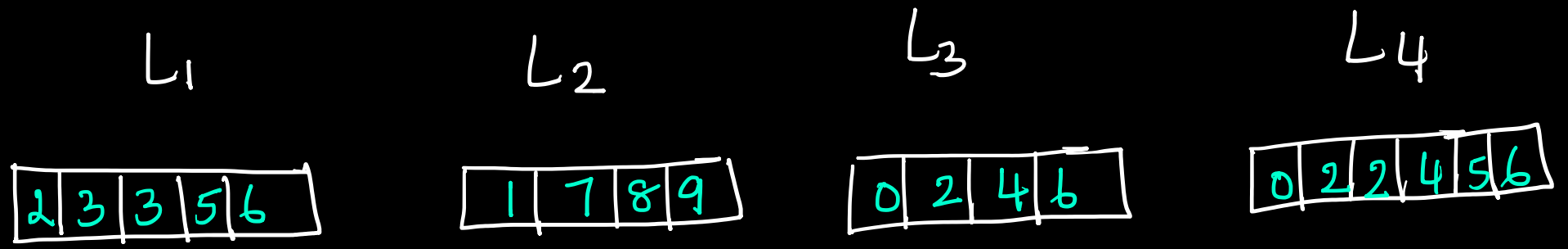
What do companies want to test?

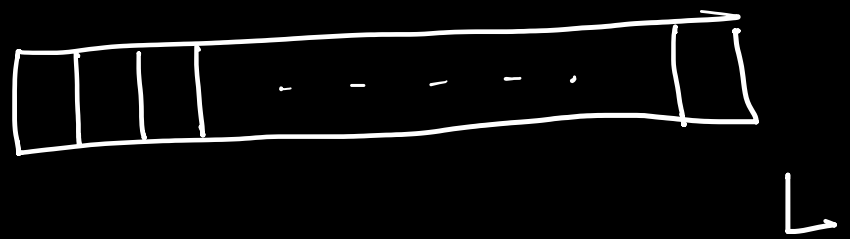
1. Depth, breadth & clarity of understanding of core concepts in Math, DS, ML, DL
2. Problem-solving
3. Ability to write code
4. Tackle open-ended problems [experienced/senior roles]

Q1 . Implement a function to merge K-sorted arrays/lists. Explain it’s time and space complexity

Microsoft, Amazon, Netflix etc.

Pseudo-code/Logic: 5-10 mins
+
Implementation : ~15 mins





Time - complx:

Space - complx:

Variations:

→ optimzed find-min [cs]

→ Merge sorted very large files [cs]

Q2 . Given a sample of observations, how would you estimate the 99^{th} percentile value of the population?

Google Data Science

Central-Limit-Theorem

Non-parametric stats:

$S = \{x_1, x_2, \ldots x_K\}$ → bootstrap →

$S_1 \xrightarrow{99^{th}\ per} p_1$

$S_2 \rightarrow p_2$

$\vdots \quad \vdots$

$S_m \rightarrow p_m$

$p_1, p_2, \ldots p_m \rightarrow \hat{p}$

c.i of p

Variations:

1. estimate median — CLT for percentiles [hardcore stats]
2. estimate uncertainty in model weights
   - Bootstrapping
   - Bayesian models.

BNN

Source

Q3 . Given the following statistics, what is the probability that a woman has cancer if she has a positive mammogram result?

- One percent of women over 50 have breast cancer.
- Ninety percent of women who have breast cancer test positive on mammograms.
- Eight percent of women will have false positives.

Most Data Science roles across almost all companies

Practice problems:

https://www.statisticshowto.com/bayes-theorem-problems/

http://wwwf.imperial.ac.uk/~atw/Bayes.pdf

http://gtribello.github.io/mathNET/bayes-theorem-problems.html

$$P(C \mid G) = 0.01$$

$$P(+ \mid C) = 0.90$$

$$P(+ \mid NC) = 0.08$$

Calc: $P(C \mid +) = \dfrac{P(C \cap +)}{P(+)}$

$$= \frac{P(+ \mid C)\ P(C)}{P(+ \cap C)\ P(+ \cap NC)}$$

$$= \frac{0.9 \times 0.01}{(0.9 \times 0.01) + (0.08 \times 0.99)}$$

$$= 0.10 = 10\%$$

Q4. Implement a function SGDLinearRegression() using basic NumPy/SciPy/Python which has the following properties:

1. Input : Dataset, lambda(multiplier to the regularization term), batch-size, #epochs
2. Implement Linear regression with Squared loss and L1-regularization
3. Implement batch-SGD optimizer
4. Output: trained Weight-vector

Pseudo-code: 5-10 mins
+
Python implementation: ~20 mins

Amazon, Netflix & many startups

$$\mathcal{L} = \min_{w} \sum_{j=1}^{n} (y_j - w^T x_j)^2 + \|w\|_1 \rightarrow |w_1| + |w_2| + \cdots |w_d|$$

$$\frac{\partial L}{\partial w_i} = 2 \cdot \sum_{j=1}^{n} (y_j - w^T x_j) \cdot (-x_{ji}) + \text{sign}(w_i)$$

$$\text{sign}(w_i) = \begin{cases} +1 & w_i > 0 \\ -1 & w_i < 0 \\ 0 & w_i = 0 \end{cases}$$

$$w_i^{new} = w_i^{old} - \eta \frac{\partial L}{\partial w_i}$$

```
η = 0.1
standardize(D)
w = random-vector(d)
for i in 1,2,... epochs
    for b in Batch(D,bs)
        for j in 1,2,...d
            wj -= η* dL-dwj(D,w,j)

Update(η)
```

Variations:

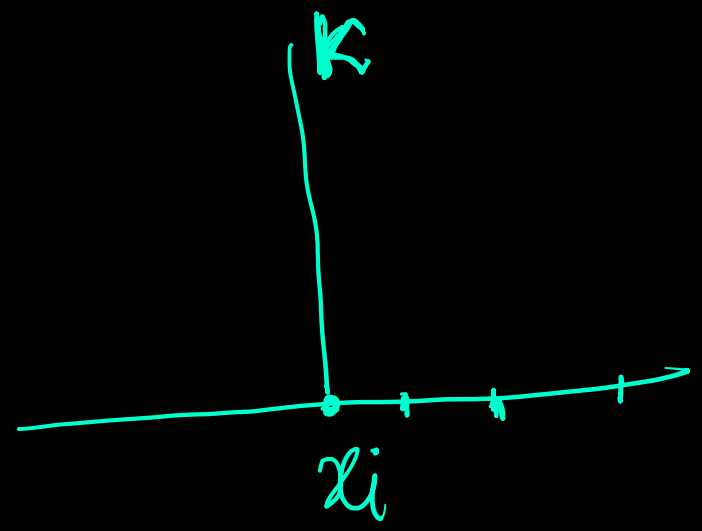
- → logistic regression or L_0-SVM
- → elastic-Net reg.
- → Custom-loss function
- → Adam optimizer instead of SGD

Q5. Why is RBF kernel the most popular & default kernel in SVMs, especially when we don’t have custom-kernels for the problem at hand

Facebook/Google ML roles

$$K(x_i, x_j) = \exp\left\{ - \frac{\|x_i - x_j\|^2}{2\sigma^2} \right\}$$

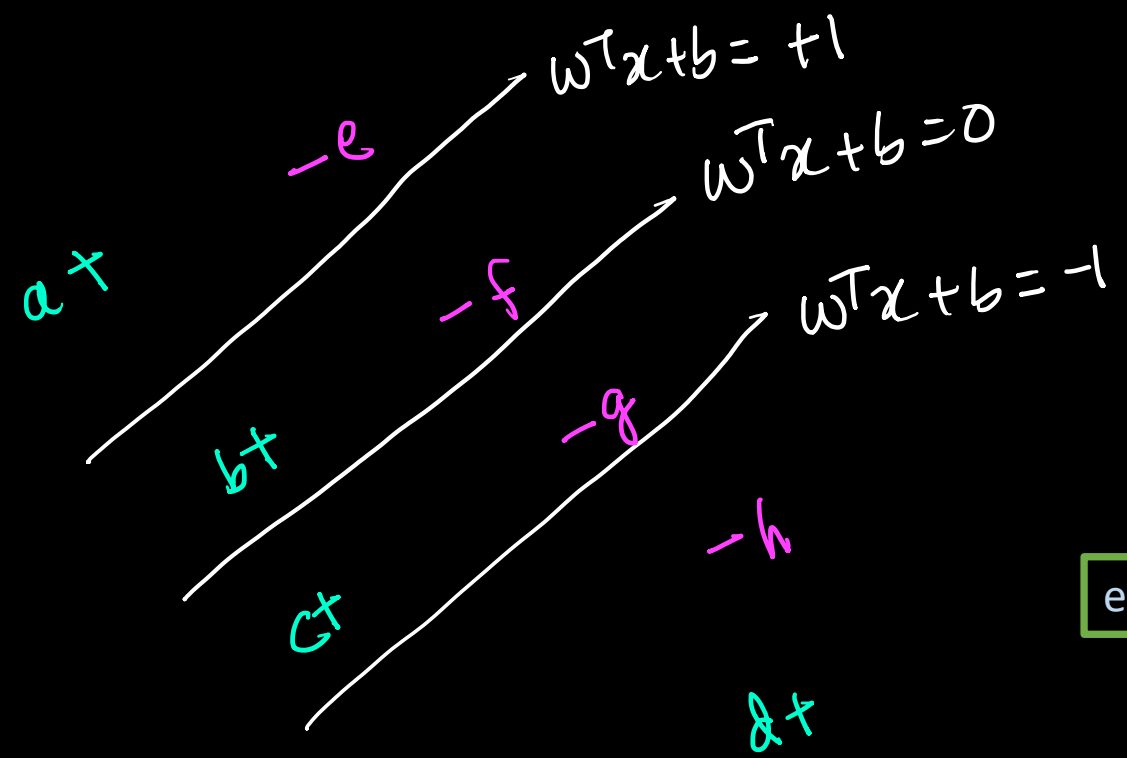
$x_i, x_j \in \mathbb{R}^d$



Variations:

1. Why do margin & regularization look similar?
2. Why are SVMs not as popular as GBDT / RF?

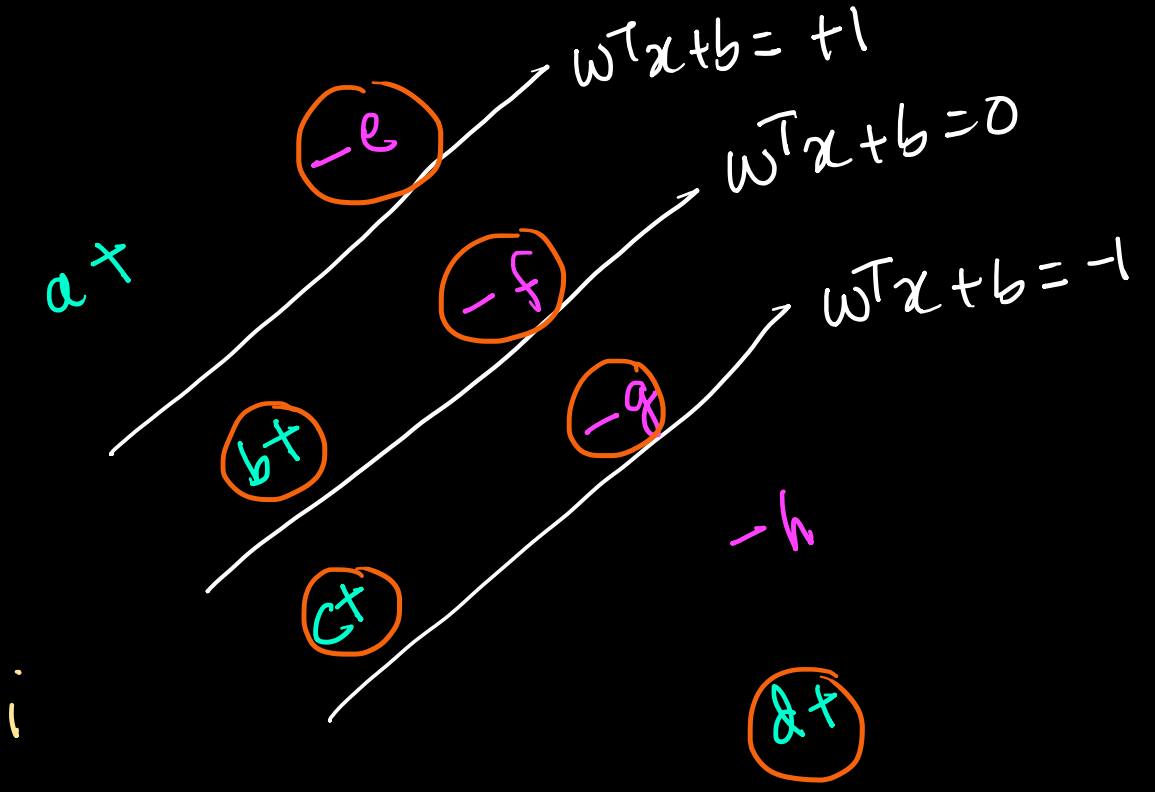
Linear SVM:



e.g: Oracle/Yahoo!

Q6. Which of these points have a non-zero slack in the Soft-margin Linear-SVM's primal formulation.

Linear SVM:



$$\min_{w, b, \xi_i} \quad \frac{\|w\|}{2} + C \sum_{i=1}^{n} \xi_i$$

s.t $\boxed{1 - y_i (w^T x_i + b)} \geq \xi_i \quad \forall i$

$\xi_i \geq 0 \quad \forall i$

$$\max\left(0, 1 - y_i (w^T x_i + b)\right)$$

Q7. While constructing GBDT, how do you compute the multiplier term for each base-learner

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x).$$

model using first 'm' base-learners

m^{th} base-learner

already known

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x).$$

model already fit on pseudo-residuals

3. Compute multiplier γ_m by solving the following one-dimensional optimization problem:

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L\left(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)\right).$$

constant

constant

Variations

1. step-by-step GBDT pseudo-code
2. modify GBDT to build k base-learners in each step.
3. Distributed GBDT [senior roles]

Q8. Explain step-by-step (or via pseudo-code), how backpropagation occurs through a batch-normalization layer?

[15-min]

e.g: Flipkart, Microsoft

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1...m}\}$;
Parameters to be learned: γ, β

Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

input

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma^2_{\mathcal{B}} \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_{\mathcal{B}})^2 \qquad \text{// mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma^2_{\mathcal{B}} + \epsilon}} \qquad \text{// normalize}$$

$$y_i \leftarrow \gamma\widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

output

Algorithm 1: Batch Normalizing Transform, applied to activation x over a mini-batch.

original paper

[Forward-pass]

Backward pass:

simple derivatives & Chain rule

https://kevinzakka.github.io/2016/09/14/batch_normalization/

https://kratzert.github.io/2016/02/12/understanding-the-gradient-flow-through-the-batch-normalization-layer.html

Variations

1. explain backprop step by step [30 min]
2. backprop with Dropout / Maxpool / ....
3. L_1 reg of weights in a layer ?

Q9. Explain the training of a GAN, preferably with pseudo-code.

[15 min]

Amazon, Google, MSR

- depends on how you set-up the problem
- multiple correct alternatives

https://www.tensorflow.org/tutorials/generative/dcgan

https://machinelearningmastery.com/how-to-code-the-generative-adversarial-network-training-algorithm-and-loss-functions/

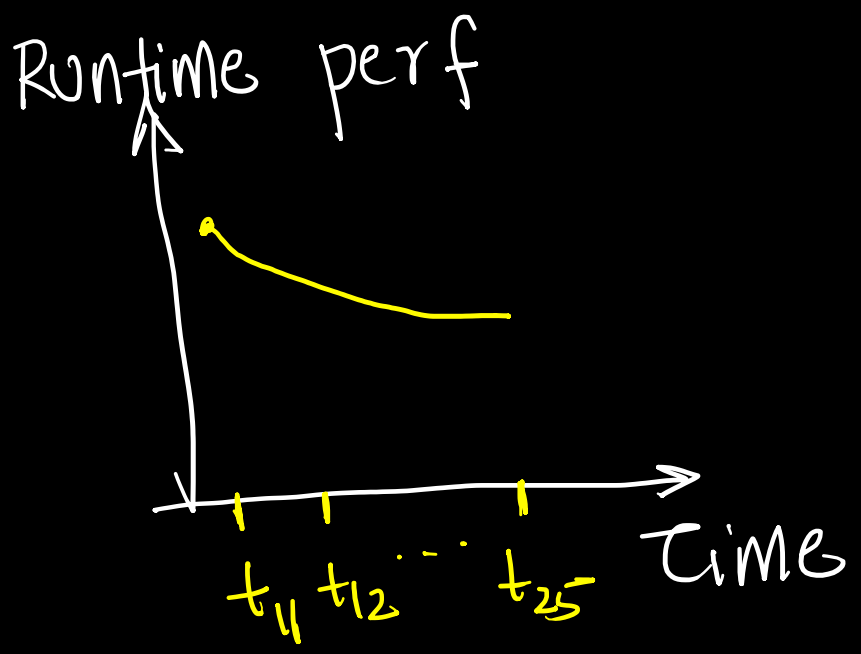
Variations

1. LSTM training

2. Transformer https://www.tensorflow.org/tutorials/text/transformer

Q10. You have an ML classification model, M, running in production trained on data from days 1 to 10. Today is day 25. What metric(s) would you keep track of, to update the production model using more recent data.

Amazon, Microsoft, Facebook



→ Runtime dataset difference across days

The hardest questions in most interviews are based around "Why?"

56 Sample Interview/Conceptual Questions with Audio explanation:

https://www.appliedaicourse.com/lecture/11/applied-machine-learning-online-course/4209/sample-interview-and-conceptual-questions-audio/69/module-10-live-sessions-on-miscellaneous-topics-audio-answers