

Support Vector Machines

(see Lagrange, primal & dual from notes book)

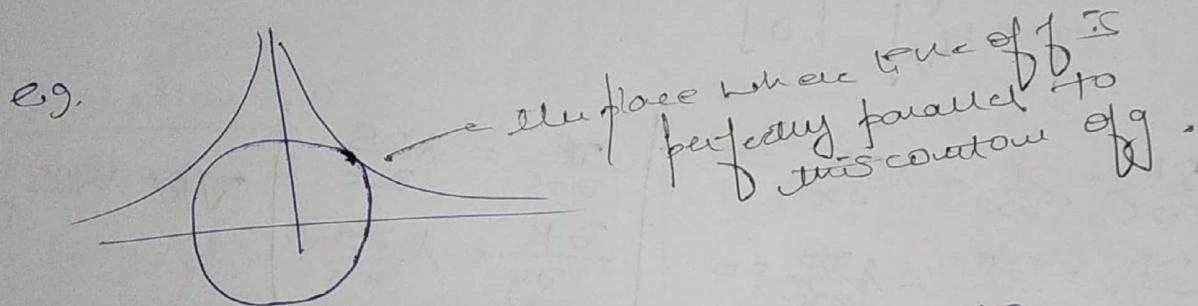
The Lagrangian \rightarrow for constrained optimiz problem

optimiz problems

constrained unconstrained

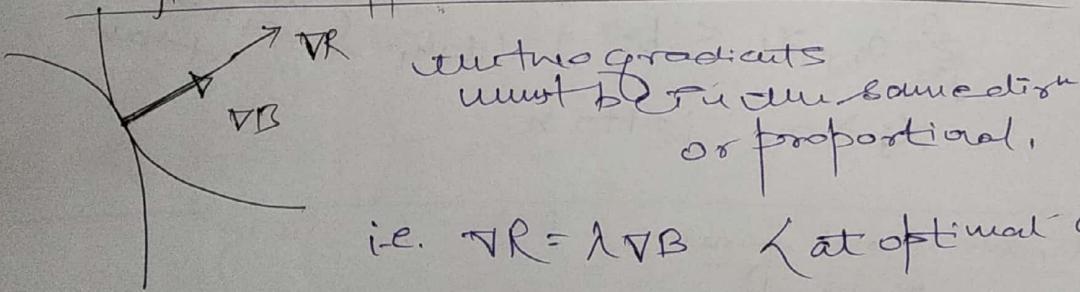
e.g. $f(\mathbf{w}, \mathbf{y}) = \frac{1}{2} \mathbf{w}^T \mathbf{w}$ } unconstrained
 $\text{max}_{\mathbf{w}, \mathbf{y}}$ $\mathbf{w}^T \mathbf{y} = \mathbf{c}$

e.g. $f(\mathbf{w}, \mathbf{y}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \mathbf{c}^T \mathbf{y}$ } constrained optimiz
 $\text{max}_{\mathbf{w}, \mathbf{y}}$ $\mathbf{w}^T \mathbf{y} = \mathbf{c}^T \mathbf{y} = \mathbf{y}$ problem
 $g(\mathbf{w}, \mathbf{y}) = \mathbf{x}^T \mathbf{y} \leq 1$ constraint



generally $f = R$ (Revenue) \rightarrow we want to maximize revenue
 $g = B$ (Budget) you have some sort of dollar limit b) Budget.

different approach to solve this problem



i.e. $\nabla R = \lambda \nabla B$ } at optimal condition

so basically Eq needs

$$\left\{ \begin{array}{l} \nabla R = \lambda \nabla B \\ A \mathbf{B}^T \mathbf{w} + b = \mathbf{c}^T \mathbf{y} = \mathbf{y} \end{array} \right.$$

gives three eq's
one more

Lagrangian is just a way to pack up these two eqns into a single entity

$$L(x, y, \lambda) = R(x, y) - \lambda (B(x, y) - b)$$

↑ ↑ ↑
 lagrangian optimiz'g fu' constraint.

Now

$$\begin{bmatrix} \nabla L = 0 \\ \frac{\partial L}{\partial x} \\ \frac{\partial L}{\partial y} \\ \frac{\partial L}{\partial \lambda} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \xrightarrow{\text{find } (x^*, y^*, \lambda^*)}$$

$$\begin{aligned} \frac{\partial L}{\partial x} &= \frac{\partial R}{\partial x} - \lambda \frac{\partial B}{\partial x} = 0 \\ \frac{\partial L}{\partial y} &= \frac{\partial R}{\partial y} - \lambda \frac{\partial B}{\partial y} = 0 \\ \frac{\partial L}{\partial \lambda} &= B(x, y) - b = 0 \end{aligned} \quad \left. \begin{aligned} \frac{\partial R}{\partial x} &= \lambda \frac{\partial B}{\partial x} \\ \frac{\partial R}{\partial y} &= \lambda \frac{\partial B}{\partial y} \end{aligned} \right\} \nabla R = \lambda \nabla B$$

Primal standard form

$$\max Z = \sum_{j=1}^n c_j x_j$$

s.t.

$$\sum_{j=1}^n a_{i,j} x_j \leq b_i \quad i=1 \text{ to } m \quad \left. \begin{array}{l} \text{w.r.t. row} \\ \text{constraint is less than} \end{array} \right\}$$

$$x_j \geq 0, \quad j=1 \text{ to } n$$

Dual standard form

$$\min W = \sum_{i=1}^m u_i b_i$$

s.t.

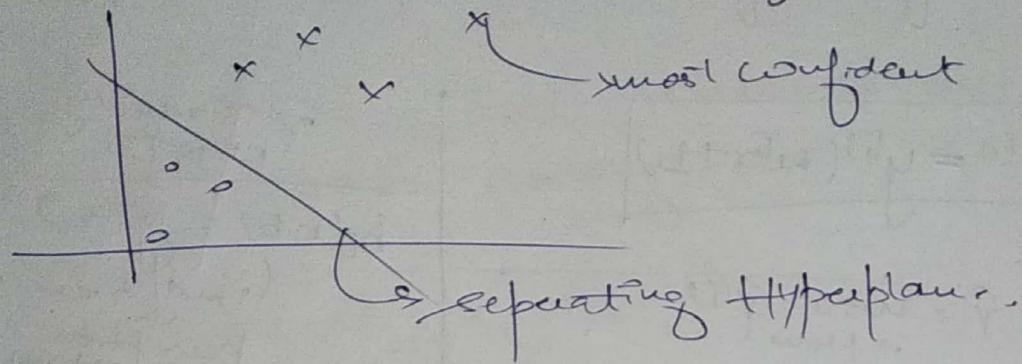
$$\sum_{j=1}^n a_{i,j} u_j \geq c_i, \quad j=1 \text{ to } n$$

$$u_i \geq 0 \quad i=1 \text{ to } m$$

both greater than

Support Vector Machines (classifer)

We want correct and confident prediction



Hyperplane

The subspace whose dimension is one less than its ambient space.

Notation

hypothesis,

$$h_{w,b}(x) = g(w^T x + b)$$

$$y \in \{-1, 1\}$$

instead of {0,1}

using perceptron algorithm

we can get directly

$$g(w^T x + b) = 1 \text{ or } -1$$

$$\text{i.e. } \begin{cases} g(z) = 1 & \text{if } z \geq 0 \\ g(z) = -1 & \text{if } z < 0 \end{cases}$$

label

$$h_0(x) = g(w^T x)$$

$$h_0(x) = p(y=1|x; w, b)$$

$$h_0(x) \geq 0.5, \theta^T x \geq 0 \quad \left\{ \begin{array}{l} y=1 \\ \text{or } \theta^T x > 0 \end{array} \right.$$

confident predict

Functional Margin

Since we want confident prediction,
we define functional margin -

for 1 TE

$$\boxed{\hat{y}^{(i)} = y^{(i)}(w^T x + b)}$$

for all TE

$$\hat{y} = \max_{i \in \text{from}} \hat{y}^{(i)}$$

z is actually
 $w^T x + b$

$$h(w, b) = g(z)$$

(but here ± 1
perception
algorithm)

$\{ \hat{y} y^{(i)} = 1$, then $w^T x + b$ must be large for
correct prediction.
(z must be large)

if $y^{(i)} = -1$ then $w^T x + b$ very large.

Not true

functional margin can only be used to check
correct prediction or not but not confidence

i.e. if

| | |
|---------------------|----------------------|
| $\hat{y}^{(i)} > 0$ | correct prediction |
| $\hat{y}^{(i)} < 0$ | incorrect prediction |

because:

$w^T x + b$ and $2w^T x + 2b$ ~~are other values of~~
functional margin but does not change
anything geometrically.

$$\text{i.e. } g(w^T x + b) = g(2w^T x + 2b)$$

for training set $S = \{(x^{(i)}, y^{(i)}); i=1\text{ to }m\}$
 We want to find
 $\boxed{\hat{y} = \min_{i=1\text{ to }m} y^{(i)}}$

* Among all TE, what is the min value of FM is
 of complete Training set.

Geometric Margin

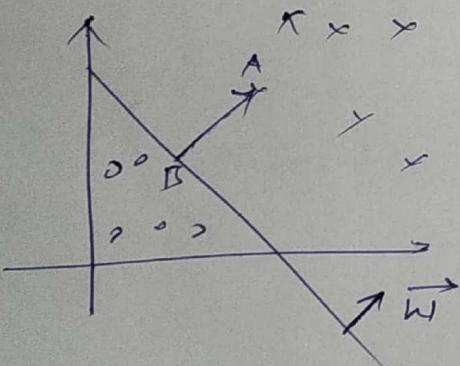
can be actually used to measure confidence.

Not a given line

but because const by change in
just off.

→ we are measuring confident classifier not a
given line.

i.e. we are trying to find a line/classifier with
max confidence or GM.



for one TE

$$\boxed{y^{(i)} = y^{(i)} \left(\left(\frac{w}{\|w\|} \right)^T x^{(i)} + \frac{b}{\|w\|} \right)}$$

for all TE

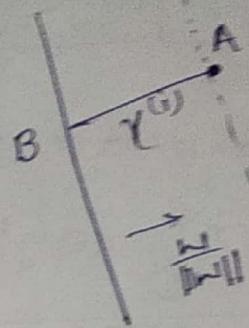
$$\boxed{\hat{y} = \min_{i=1\text{ to }m} y^{(i)}}$$

Derivation of GM \rightarrow decision boundary

Let point A is $x^{(i)}$
unit vector \hat{w} to $wB = \frac{w}{\|w\|}$

Vector ~~to~~ $AB = \gamma^{(i)} \frac{w}{\|w\|}$

$$\text{So } BS = x^{(i)} - \gamma^{(i)} \frac{w}{\|w\|}$$



It lies on PB

so, satisfy by $w^T b + b = 0$

$$\Rightarrow w^T \left(x^{(i)} - \gamma^{(i)} \frac{w}{\|w\|} \right) + b = 0$$

$$\Rightarrow \boxed{\gamma^{(i)} = \frac{w^T x^{(i)} + b}{\|w\|}} = \frac{\gamma^{(i)}}{\|w\|}$$

Note

$$GM(\gamma) \leq PM(\hat{\gamma})$$

$$\gamma^{(i)} \leq \hat{\gamma}^{(i)} \quad i=1 \text{ to } m$$

Optimizer problem (Simplification)

$$\textcircled{1} \quad \begin{array}{ll} \max_{\gamma, w, b} & \gamma \\ \text{s.t.} & y^{(i)}(w^T x^{(i)} + b) \geq \gamma, \quad i=1 \dots m \\ & \|w\| = 1 \end{array} \quad \left| \begin{array}{l} \text{max of GM} \\ \text{GMTE} \geq GM \\ \text{so} \\ \|w\| = 1 \end{array} \right.$$

In the constraint $\|w\| = 1$ is
non-convex
(not possible to get a
global unique optimum)

$$\textcircled{2} \quad \begin{array}{ll} \max_{\gamma, w, b} & \frac{\gamma}{\|w\|} \\ \text{s.t.} & y^{(i)}(w^T x^{(i)} + b) \geq \gamma, \quad i=1 \dots m \end{array} \quad \left| \begin{array}{l} \text{max of GM} \\ \text{FMTE} \geq FM \end{array} \right.$$

In the objective function $\frac{\gamma}{\|w\|}$ is non-convex.

Note: Since we can scale w in any way, we will always have $\|w\| = 1$.

$$\textcircled{3} \quad \begin{array}{ll} \max_{w, b} & \frac{1}{\|w\|} \\ \text{s.t.} & y^{(i)}(w^T x^{(i)} + b) \geq 1, \quad i=1 \dots m. \end{array}$$

Also

$$\boxed{\begin{array}{ll} \max_{w, b} & \frac{1}{2} \|w\|^2 \\ \text{s.t.} & y^{(i)}(w^T x^{(i)} + b) \geq 1 \end{array}} \quad \left| \begin{array}{l} \text{convex quadratic} \\ \text{objective func} \\ \text{linear constraint} \\ \text{its soln gives us optimal} \\ \text{margin classifier} \\ \rightarrow \text{Not in primal form} \end{array} \right.$$

Optimization problem using lagrangian dual form

Benefits -

→ Used to show SVM depends on very small no. of support vectors

ii) The optimization problem can be expressed in terms of inner product only.

$$\langle \mathbf{x}^{(i)}, \mathbf{x}^{(j)} \rangle \text{ or } (\mathbf{x}^{(i)})^T \mathbf{x}^{(j)}$$

{ for KKT }
no need to find ϕ as going high dimension

Lagrange duality → solving constrained optimization problem

$$\min_w f(w)$$

$$\text{s.t. } h_i(w) = 0, i=1 \dots d$$

Solving it using Lagrange multiplier method.

$$L(w, \beta) = f(w) + \sum_{i=1}^d \beta_i h_i(w)$$

Lagrange multiplier

Now we have to find PD of L to zero.

$$\boxed{\frac{\partial L}{\partial w_i} = 0, \frac{\partial L}{\partial \beta_i} = 0}$$

and solve for w and β

Generalizing primal optimization problems to dual optimization problem -

$$\min_w f(w)$$

$$\text{st. } \begin{cases} g_i(w) \leq 0 & , i = 1 \dots k \\ h_i(w) = 0 & , i = 1 \dots l \end{cases} \quad \left. \begin{array}{l} \text{primal constraints} \\ \text{equality} \end{array} \right\}$$

Generalized Lagrangian

$$L(w, \alpha, \beta) = -f(w) + \sum_{i=1}^k \alpha_i g_i(w) + \sum_{i=1}^l \beta_i h_i(w)$$

Lagrangian multipliers.

Maximize the Lagrangian w.r.t. α, β

consider, Op where p stands for primal.

$$\text{Op}(w) = \max_{\alpha, \beta: \alpha_i \geq 0} L(w, \alpha, \beta)$$

$$\text{Op}(w) = \max_{\alpha, \beta: \alpha_i \geq 0} f(w) + \sum_{i=1}^k \alpha_i g_i(w) + \sum_{i=1}^l \beta_i h_i(w)$$

$$\text{Op}(w) = \begin{cases} f(w), & \text{if } w \text{ satisfies primal constraints} \\ \infty, & \text{otherwise} \end{cases}$$

If w satisfies primal constraints —

$$\boxed{\min_w f(w) = \min_w \text{Op}(w) = \min_w \max_{\alpha, \beta: \alpha_i \geq 0} L(w, \alpha, \beta)}$$

Let $p^* = \min_w \text{Op}(w)$ value of the primal problem.

Dual optimization problem (its transpose)

$$O_D(\alpha, \beta) = \min_u L(u, \alpha, \beta)$$

$$\max_{\alpha, \beta; \alpha_i \geq 0} O_D(\alpha, \beta) = \max_{\alpha, \beta; \alpha_i \geq 0} \min_u L(u, \alpha, \beta)$$

Order of maximization
in dual = max over multipliers α, β

$$\text{Let } d^* = \max_{\alpha, \beta; \alpha_i \geq 0} \min_u L(u, \alpha, \beta) \text{ called } O_D^{(d^*)}$$

$\therefore \text{max}_{\alpha, \beta} \leq \text{max}_{\alpha, \beta, u}$ of d^*

$$d^* \leq p^*$$

$$\downarrow$$

$$\max_{\alpha, \beta; \alpha_i \geq 0} \min_u L(u, \alpha, \beta) \leq \min_u \max_{\alpha, \beta; \alpha_i \geq 0} L(u, \alpha, \beta).$$

\therefore there must be optimality for equality condⁿ
 u^*, α^*, β^*

$$\text{s.t. } d^* = p^* = L(u^*, \alpha^*, \beta^*)$$

at equality condⁿ it must satisfy KKT
 (Karush-Kuhn-Tucker) conditions \rightarrow

KKT (Karush-Kuhn-Tucker) conditions

$$\frac{\partial}{\partial w_i} L(w^*, \alpha^*, \beta^*) = 0 \quad i=1 \dots n$$

$$\frac{\partial}{\partial \beta_i} L(w^*, \alpha^*, \beta^*) = 0 \quad i=1 \dots L$$

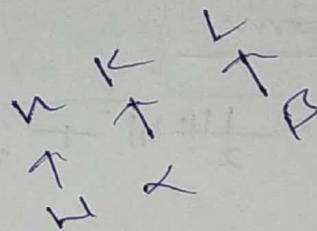
$$\alpha_i g_i(w^*) = 0$$

$$\alpha_i \geq 0$$

$$g_i(w^*) \leq 0$$

$\Rightarrow \alpha_i \neq 0 \quad \left\{ \begin{array}{l} i=1 \dots K \\ i=1 \dots n \end{array} \right.$

(used for support vectors)



w

α

β

Optimal Margin classifier

Primal form

Optimization problem standard primal form -

$$\underset{w, b}{\text{min}} \frac{1}{2} \|w\|^2$$

$$\text{s.t. } -y^{(i)}(w^T x^{(i)} + b) + 1 \leq 0 \quad \curvearrowright g_i(w)$$

Dual form

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^m \alpha_i [1 - y^{(i)}(w^T x^{(i)} + b)]$$

$$= \frac{1}{2} \|w\|^2 - \sum_{i=1}^m \alpha_i [y^{(i)}(w^T x^{(i)} + b) - 1]$$

$\therefore \alpha_i > 0$ if $g_i(w^*) = 0$

so only the ones for which functional margin is 1

$$\alpha_i = 1$$

for others $\alpha_i = 0$

support vectors

(less calculation
does not consider
other TEs for
PB calculation)

for showing support vectors needed just transform

from dual form and
show KKT condition $\nabla L(\alpha) = 0$

(11)

Representation of lagrangian only using
of inner product (Kernel trick)

$$L(u, b, \alpha) = \frac{1}{2} \|u\|^2 - \sum_{i=1}^m \alpha_i [y^{(i)}(u^T x^{(i)}) - 1]$$

Dual = maximum
con. for w.r.t. u, b take gradient

$$\nabla_u L(u, b, \alpha) = \frac{\partial}{\partial u} L(u, b, \alpha) \\ = \|u\| - \sum_{i=1}^m \alpha_i y^{(i)} x^{(i)} = 0$$

so,

$$u = \sum_{i=1}^m \alpha_i y^{(i)} x^{(i)}$$

$$u^T = \sum_{i=1}^m \alpha_i y^{(i)} (x^{(i)})^T$$

and $\frac{\partial L}{\partial b} = - \sum_{i=1}^m \alpha_i y^{(i)} = 0$

Substitute in ① to get other form

$$L(u, b, \alpha) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y^{(i)} y^{(j)} (x^{(i)})^T + \sum_{i=1}^m \alpha_i - \sum_{i=1}^m \sum_{j=1}^m \alpha_i y^{(i)} x^{(i)} \alpha_j y^{(j)} (x^{(j)})^T - b \sum_{i=1}^m \alpha_i$$

$$L(u, b, \alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y^{(i)} y^{(j)} (x^{(i)})^T (x^{(j)})^T$$

$$L(u, b, \alpha) = \max_{\alpha} \Theta_2(u, b) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m y^{(i)} y^{(j)} \alpha_i \alpha_j \langle x^{(i)}, x^{(j)} \rangle$$

$$\alpha_i \geq 0, i=1 \dots n$$

$$\sum_{i=1}^m \alpha_i y^{(i)} = 0$$

Only express and
values of inner
product of inner
product of $x^{(i)}$
and $x^{(j)}$.
When we have to
use $\Phi(x^{(i)})$,
 $\Phi(x^{(j)})$
we can directly use
 $K(x^{(i)}, x^{(j)}) = \Phi(x^{(i)})^T \Phi(x^{(j)})$

For given Example \rightarrow We will predict if
 $w^T x + b > 0$

$$\begin{aligned} w^T x + b &= \left(\sum_{i=1}^m \alpha_i y^{(i)} \phi^{(i)} \right)^T x + b \\ &= \sum_{i=1}^m \alpha_i y^{(i)} \phi^{(i)} \cdot \phi^{(i)}^T x + b \\ &= \sum_{i=1}^m \alpha_i y^{(i)} \langle \phi^{(i)}, x \rangle + b \end{aligned}$$

against any function
we can use kernel
dot product.

$$b^* = -\frac{\sum_{i:y^{(i)}=-1} \alpha_i w^T \phi^{(i)} + \sum_{i:y^{(i)}=1} \alpha_i w^T \phi^{(i)}}{2}$$

Kernel Trick

We can use many feature maps $\phi(x)$ from attributes to features

e.g. $x \rightarrow \text{say } x^*$

In that case inner product in u & v will be replaced by $\langle \phi(x_i^*), \phi(x_j^*) \rangle$.

Note: we define kernel

$$k(x, z) = \phi(x)^T \phi(z)$$

$$k(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$

↳ effective way to calculate $k(x, z)$ can get it simply to learn in the high dimension feature space given by ϕ

without even having to explicitly find or represent vector $\phi(x)$.

Types of Kernels

Use linear kernel if this linearly separable or features is very large.

as others go in high dimension until linearly and computationally very expensive can overfit also.

funda →

generalization linear better.

i) if $m > n$
linearly if linearly separable
try others to do

ii) if $n > m$
use linear (try to do)

Linear.

RBF

Gaussian

tanh

Majority voting logic
in support vector
occurs because
is the label of
the point / the
say category

One vs one SVM

finds the
or classifies in category to which it is maximum in OVR SVM

(and) classifies

For two classes of data points
1st
2nd
3rd
4th
5th
6th
7th
8th
9th
10th
11th
12th
13th
14th
15th
16th
17th
18th
19th
20th
21st
22nd
23rd
24th
25th
26th
27th
28th
29th
30th
31st
32nd
33rd
34th
35th
36th
37th
38th
39th
40th
41st
42nd
43rd
44th
45th
46th
47th
48th
49th
50th
51st
52nd
53rd
54th
55th
56th
57th
58th
59th
60th
61st
62nd
63rd
64th
65th
66th
67th
68th
69th
70th
71st
72nd
73rd
74th
75th
76th
77th
78th
79th
80th
81st
82nd
83rd
84th
85th
86th
87th
88th
89th
90th
91st
92nd
93rd
94th
95th
96th
97th
98th
99th
100th
101st
102nd
103rd
104th
105th
106th
107th
108th
109th
110th
111th
112th
113th
114th
115th
116th
117th
118th
119th
120th
121st
122nd
123rd
124th
125th
126th
127th
128th
129th
130th
131st
132nd
133rd
134th
135th
136th
137th
138th
139th
140th
141st
142nd
143rd
144th
145th
146th
147th
148th
149th
150th
151st
152nd
153rd
154th
155th
156th
157th
158th
159th
160th
161st
162nd
163rd
164th
165th
166th
167th
168th
169th
170th
171st
172nd
173rd
174th
175th
176th
177th
178th
179th
180th
181st
182nd
183rd
184th
185th
186th
187th
188th
189th
190th
191st
192nd
193rd
194th
195th
196th
197th
198th
199th
200th
201st
202nd
203rd
204th
205th
206th
207th
208th
209th
210th
211st
212nd
213rd
214th
215th
216th
217th
218th
219th
220th
221st
222nd
223rd
224th
225th
226th
227th
228th
229th
230th
231st
232nd
233rd
234th
235th
236th
237th
238th
239th
240th
241st
242nd
243rd
244th
245th
246th
247th
248th
249th
250th
251st
252nd
253rd
254th
255th
256th
257th
258th
259th
260th
261st
262nd
263rd
264th
265th
266th
267th
268th
269th
270th
271st
272nd
273rd
274th
275th
276th
277th
278th
279th
280th
281st
282nd
283rd
284th
285th
286th
287th
288th
289th
290th
291st
292nd
293rd
294th
295th
296th
297th
298th
299th
300th
301st
302nd
303rd
304th
305th
306th
307th
308th
309th
310th
311st
312nd
313rd
314th
315th
316th
317th
318th
319th
320th
321st
322nd
323rd
324th
325th
326th
327th
328th
329th
330th
331st
332nd
333rd
334th
335th
336th
337th
338th
339th
340th
341st
342nd
343rd
344th
345th
346th
347th
348th
349th
350th
351st
352nd
353rd
354th
355th
356th
357th
358th
359th
360th
361st
362nd
363rd
364th
365th
366th
367th
368th
369th
370th
371st
372nd
373rd
374th
375th
376th
377th
378th
379th
380th
381st
382nd
383rd
384th
385th
386th
387th
388th
389th
390th
391st
392nd
393rd
394th
395th
396th
397th
398th
399th
400th
401st
402nd
403rd
404th
405th
406th
407th
408th
409th
410th
411st
412nd
413rd
414th
415th
416th
417th
418th
419th
420th
421st
422nd
423rd
424th
425th
426th
427th
428th
429th
430th
431st
432nd
433rd
434th
435th
436th
437th
438th
439th
440th
441st
442nd
443rd
444th
445th
446th
447th
448th
449th
450th
451st
452nd
453rd
454th
455th
456th
457th
458th
459th
460th
461st
462nd
463rd
464th
465th
466th
467th
468th
469th
470th
471st
472nd
473rd
474th
475th
476th
477th
478th
479th
480th
481st
482nd
483rd
484th
485th
486th
487th
488th
489th
490th
491st
492nd
493rd
494th
495th
496th
497th
498th
499th
500th
501st
502nd
503rd
504th
505th
506th
507th
508th
509th
510th
511st
512nd
513rd
514th
515th
516th
517th
518th
519th
520th
521st
522nd
523rd
524th
525th
526th
527th
528th
529th
530th
531st
532nd
533rd
534th
535th
536th
537th
538th
539th
540th
541st
542nd
543rd
544th
545th
546th
547th
548th
549th
550th
551st
552nd
553rd
554th
555th
556th
557th
558th
559th
560th
561st
562nd
563rd
564th
565th
566th
567th
568th
569th
570th
571st
572nd
573rd
574th
575th
576th
577th
578th
579th
580th
581st
582nd
583rd
584th
585th
586th
587th
588th
589th
590th
591st
592nd
593rd
594th
595th
596th
597th
598th
599th
600th
601st
602nd
603rd
604th
605th
606th
607th
608th
609th
610th
611st
612nd
613rd
614th
615th
616th
617th
618th
619th
620th
621st
622nd
623rd
624th
625th
626th
627th
628th
629th
630th
631st
632nd
633rd
634th
635th
636th
637th
638th
639th
640th
641st
642nd
643rd
644th
645th
646th
647th
648th
649th
650th
651st
652nd
653rd
654th
655th
656th
657th
658th
659th
660th
661st
662nd
663rd
664th
665th
666th
667th
668th
669th
670th
671st
672nd
673rd
674th
675th
676th
677th
678th
679th
680th
681st
682nd
683rd
684th
685th
686th
687th
688th
689th
690th
691st
692nd
693rd
694th
695th
696th
697th
698th
699th
700th
701st
702nd
703rd
704th
705th
706th
707th
708th
709th
710th
711st
712nd
713rd
714th
715th
716th
717th
718th
719th
720th
721st
722nd
723rd
724th
725th
726th
727th
728th
729th
730th
731st
732nd
733rd
734th
735th
736th
737th
738th
739th
740th
741st
742nd
743rd
744th
745th
746th
747th
748th
749th
750th
751st
752nd
753rd
754th
755th
756th
757th
758th
759th
760th
761st
762nd
763rd
764th
765th
766th
767th
768th
769th
770th
771st
772nd
773rd
774th
775th
776th
777th
778th
779th
780th
781st
782nd
783rd
784th
785th
786th
787th
788th
789th
790th
791st
792nd
793rd
794th
795th
796th
797th
798th
799th
800th
801st
802nd
803rd
804th
805th
806th
807th
808th
809th
810th
811st
812nd
813rd
814th
815th
816th
817th
818th
819th
820th
821st
822nd
823rd
824th
825th
826th
827th
828th
829th
830th
831st
832nd
833rd
834th
835th
836th
837th
838th
839th
840th
841st
842nd
843rd
844th
845th
846th
847th
848th
849th
850th
851st
852nd
853rd
854th
855th
856th
857th
858th
859th
860th
861st
862nd
863rd
864th
865th
866th
867th
868th
869th
870th
871st
872nd
873rd
874th
875th
876th
877th
878th
879th
880th
881st
882nd
883rd
884th
885th
886th
887th
888th
889th
890th
891st
892nd
893rd
894th
895th
896th
897th
898th
899th
900th
901st
902nd
903rd
904th
905th
906th
907th
908th
909th
910th
911st
912nd
913rd
914th
915th
916th
917th
918th
919th
920th
921st
922nd
923rd
924th
925th
926th
927th
928th
929th
930th
931st
932nd
933rd
934th
935th
936th
937th
938th
939th
940th
941st
942nd
943rd
944th
945th
946th
947th
948th
949th
950th
951st
952nd
953rd
954th
955th
956th
957th
958th
959th
960th
961st
962nd
963rd
964th
965th
966th
967th
968th
969th
970th
971st
972nd
973rd
974th
975th
976th
977th
978th
979th
980th
981st
982nd
983rd
984th
985th
986th
987th
988th
989th
990th
991st
992nd
993rd
994th
995th
996th
997th
998th
999th
1000th
1001st
1002nd
1003rd
1004th
1005th
1006th
1007th
1008th
1009th
10010th
10011th
10012th
10013th
10014th
10015th
10016th
10017th
10018th
10019th
10020th
10021st
10022nd
10023rd
10024th
10025th
10026th
10027th
10028th
10029th
10030th
10031st
10032nd
10033rd
10034th
10035th
10036th
10037th
10038th
10039th
10040th
10041st
10042nd
10043rd
10044th
10045th
10046th
10047th
10048th
10049th
10050th
10051st
10052nd
10053rd
10054th
10055th
10056th
10057th
10058th
10059th
10060th
10061st
10062nd
10063rd
10064th
10065th
10066th
10067th
10068th
10069th
10070th
10071st
10072nd
10073rd
10074th
10075th
10076th
10077th
10078th
10079th
10080th
10081st
10082nd
10083rd
10084th
10085th
10086th
10087th
10088th
10089th
10090th
10091st
10092nd
10093rd
10094th
10095th
10096th
10097th
10098th
10099th
100100th
100101st
100102nd
100103rd
100104th
100105th
100106th
100107th
100108th
100109th
100110th
100111st
100112nd
100113rd
100114th
100115th
100116th
100117th
100118th
100119th
100120th
100121st
100122nd
100123rd
100124th
100125th
100126th
100127th
100128th
100129th
100130th
100131st
100132nd
100133rd
100134th
100135th
100136th
100137th
100138th
100139th
100140th
100141st
100142nd
100143rd
100144th
100145th
100146th
100147th
100148th
100149th
100150th
100151st
100152nd
100153rd
100154th
100155th
100156th
100157th
100158th
100159th
100160th
100161st
100162nd
100163rd
100164th
100165th
100166th
100167th
100168th
100169th
100170th
100171st
100172nd
100173rd
100174th
100175th
100176th
100177th
100178th
100179th
100180th
100181st
100182nd
100183rd
100184th
100185th
100186th
100187th
100188th
100189th
100190th
100191st
100192nd
100193rd
100194th
100195th
100196th
100197th
100198th
100199th
100200th
100201st
100202nd
100203rd
100204th
100205th
100206th
100207th
100208th
100209th
100210th
100211st
100212nd
100213rd
100214th
100215th
100216th
100217th
100218th
100219th
100220th
100221st
100222nd
100223rd
100224th
100225th
100226th
100227th
100228th
100229th
100230th
100231st
100232nd
100233rd
100234th
100235th
100236th
100237th
100238th
100239th
100240th
100241st
100242nd
100243rd
100244th
100245th
100246th
100247th
100248th
100249th
100250th
100251st
100252nd
100253rd
100254th
100255th
100256th
100257th
100258th
100259th
100260th
100261st
100262nd
100263rd
100264th
100265th
100266th
100267th
100268th
100269th
100270th
100271st
100272nd
100273rd
100274th
100275th
100276th
100277th
100278th
100279th
100280th
100281st
100282nd
100283rd
100284th
100285th
100286th
100287th
100288th
100289th
100290th
100291st
100292nd
100293rd
100294th
100295th
100296th
100297th
100298th
100299th
100300th
100301st
100302nd
100303rd
100304th
100305th
100306th
100307th
100308th
100309th
100310th
100311st
100312nd
100313rd
100314th
100315th
100316th
100317th
100318th
100319th
100320th
100321st
100322nd
100323rd
100324th
100325th
100326th
100327th
100328th
100329th
100330th
100331st
100332nd
100333rd
100334th
100335th
100336th
100337th
100338th
100339th
100340th
100341st
100342nd
100343rd
100344th
100345th
100346th
100347th
100348th
100349th
100350th
100351st
100352nd
100353rd
100354th
100355th
100356th
100357th
100358th
100359th
100360th
100361st
100362nd
100363rd
100364th
100365th
100366th
100367th
100368th
100369th
100370th
100371st
100372nd
100373rd
100374th
100375th
100376th
100377th
100378th
100379th
100380th
100381st
100382nd
100383rd
100384th
100385th
100386th
100387th
100388th
100389th
100390th
100391st
100392nd
100393rd
100394th
100395th
100396th
100397th
100398th
100399th
100400th
100401st
100402nd
100403rd
100404th
100405th
100406th
100407th
100408th
100409th
100410th
100411st
100412nd
100413rd
100414th
100415th
100416th
100417th
100418th
100419th
100420th
100421st
100422nd
100423rd
100424th
100425th
100426th
100427th
100428th
100429th
100430th
100431st
100432nd
100433rd
100434th
100435th
100436th
100437th
100438th
100439th
100440th
100441st
100442nd
100443rd
100444th
100445th
100446th
100447th
100448th
100449th
100450th
100451st
100452nd
100453rd
100454th
100455th
100456th
100457th
100458th
100459th
100460th
100461st
100462nd
100463rd
100464th
100465th
100466th
100467th
100468th
100469th
100470th
100471st
100472nd
100473rd
100474th
100475th
100476th
100477th
100478th
100479th
100480th
100481st
100482nd
100483rd
100484th
100485th
100486th
100487th
100488th
100489th
100490th
100491st
100492nd
100493rd
100494th
100495th
100496th
100497th
100498th
100499th
100500th
100501st
100502nd
100503rd
100504th
100505th
100506th
100507th
100508th
100509th
100510th
100511st
100512nd
100513rd
100514th
100515th
100516th
100517th
100518th
100519th
100520th
100521st
100522nd<br

Error Metrics → F₁ score is used for misclassified errors.

$$\text{Precision}_j = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{for class } j$$

$$\text{Recall}_j = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{for class } j$$

$$F_{ij} = \frac{2 \times \text{Precision}_j \times \text{Recall}_j}{\text{Precision}_j + \text{Recall}_j}$$

We got 96% on test data
and 99% on training data.

Cost function.

$$J = \frac{1}{m} \sum_{i=1}^m \max(0, 1 - g_i(h^T x^{(i)} + b)) + \frac{1}{2} \|H\|^2$$

↓
regularization

* for OVS all forms e.g.
Answers to make prediction
pick i that maximizes
 $\max_i g_i(x)$